

1-1-2012

# Peer-To-Peer Multimedia And Information Sharing In Social Networking Web Sites

Shahin Talaei  
*Ryerson University*

Follow this and additional works at: <http://digitalcommons.ryerson.ca/dissertations>



Part of the [Computer Sciences Commons](#)

---

## Recommended Citation

Talaei, Shahin, "Peer-To-Peer Multimedia And Information Sharing In Social Networking Web Sites" (2012). *Theses and dissertations*. Paper 1723.

This Thesis is brought to you for free and open access by Digital Commons @ Ryerson. It has been accepted for inclusion in Theses and dissertations by an authorized administrator of Digital Commons @ Ryerson. For more information, please contact [bcameron@ryerson.ca](mailto:bcameron@ryerson.ca).

# **PEER-TO-PEER MULTIMEDIA AND INFORMATION SHARING IN SOCIAL NETWORKING WEB SITES**

By

Shahin Talaei

B.Sc. Electrical Engineering and Informatics, Budapest University of Technology  
and Economics, 2007

A thesis

Presented to Ryerson University

in partial fulfillment of the

requirement for the degree of

Master of Science

in the Program of

Computer Science

Toronto, Ontario, Canada, 2012

© Shahin Talaei 2012

## **Author's Declaration**

### AUTHOR'S DECLARATION FOR ELECTRONIC SUBMISSION OF A THESIS

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I authorize Ryerson University to lend this thesis to other institutions or individuals for the purpose of scholarly research

I further authorize Ryerson University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

I understand that my thesis may be made electronically available to the public.

# PEER-TO-PEER MULTIMEDIA AND INFORMATION SHARING IN SOCIAL NETWORKING WEB SITES

Shahin Talaei

M.Sc. Computer Science, Ryerson University, 2012

## **Abstract**

This thesis examines the performance for multimedia distribution and information sharing of social-networking web sites, with a focus on user networks in Facebook. We used real user network data from Facebook together with a synthetic Facebook network in the performance-testing experiments. We tested performance for multimedia distribution and information sharing using three different types of overlay networks: Facebook; structured peer-to-peer (ring topology); and unstructured peer-to-peer (mesh topology). The experiments used Network Simulator 2 (Ns-2) to simulate the network topologies. The results show that structured Peer-to-Peer has the best performance in terms of information transfer, and Facebook has the best performance in regards to average throughput. This thesis shows the strengths and weaknesses of online social networking while sharing information and multimedia content.

## **Acknowledgments**

I would like to express my gratitude to my supervisor, Dr. Abdolreza Abhari, for all his dedicated hard work. It is thanks to his continuous support and encouragement that this thesis is accomplished.

I am grateful to the committee Dr. Sadeghian, Dr. Ferworn and Dr. Ding for reading and providing feedback on my work.

I am indebted to my friends Subir Biswas, Graham Holker and Fouad Butt. I am proud of working with them, and being their fellow researcher. I want to thank Amir Pasha Pashandi for encouraging me to apply for the program. I also would like to thank the following friends who contributed to my time at Ryerson University: Fatemeh Pakzad, Abdolkarim Dabagh, Shermineh Ghasemi, Shilpi Verma, Sonal Patel, Dr. Maryam Davoudpour, Dr. Mehrdad Tirandazian, Maryam Bashardoust, Saeed Rashwand, Kian Farsandaj, Touhid Ghasemi, Vahab Akbarzadeh, and Masoud Zamani.

It is a pleasure to sincerely thank the administrative staff members Mrs. Lucia Flaim, Mrs. Lori Haddad and Deborah Mckay for their tolerance and extended help whenever needed.

I would like to extend my gratitude to William Zereneh, Ivan Rubiales and Mohammad Fayaz for their honest help in technical support.

I just cannot stop appreciating my family for their continuous love and support.

## **Dedication**

*To my Family*

## Table of Contents

1. Introduction .....	1
1.1 Information Sharing in Web 2.0.....	1
1.2 Social Networking Web Sites .....	2
1.3 Employing Social Networking in P2P Networks .....	3
1.4 Thesis Motivation.....	3
1.5 Thesis Objectives and Scope.....	5
1.6 Research Contributions .....	6
1.7 Thesis Outline .....	6
2. Background Information and Related Works .....	7
2.1 Background Information .....	7
2.1.1 Facebook.....	7
2.1.2 Other Online Social Networks:.....	8
2.1.3 Small World Phenomena .....	8
2.1.4 Spotify .....	9
2.1.5 YouTube .....	11
2.1.6 Overlay Networks .....	12
2.1.6.1 Two of the most popular approaches to constructing an overlay network .....	13
2.1.7 P2P Networks .....	13

2.1.7.1 Chord .....	16
2.1.7.2 Classification of P2P applications .....	16
2.1.7.3 P2P Simulators .....	17
2.1.7.4 BitTorrent .....	17
2.1.7.5 Distributed Hash Tables .....	20
2.1.7.6 Group Communication .....	21
2.1.8 Ring .....	22
2.1.9 Network Simulator 2(Ns-2) .....	23
2.2 Related Works .....	25
2.2.1 Hybrid P2P Systems .....	25
2.2.2 Multimedia Characterization in Social Networking Web 2.0 Sites.....	25
2.2.3 Analyzing YouTube and Facebook .....	26
2.2.4 Multimedia sharing in overlay networks .....	27
3. Proposed Model and Methodology.....	28
3.1 Integration of Social Networking with P2P Network .....	28
3.2 Simulating Facebook Network.....	29
3.3 Simulating P2P Networks .....	33
3.4.1 Simulation Architecture.....	33
3.4 Performance Metrics .....	36
3.5 Summary .....	37



4. Experiments and Results .....	38
4.1 Creating Facebook, Peer-to-Peer and Ring Overlay Networks.....	38
4.1.1 Simulation of Simple Information Transfer .....	39
4.1.2 Short Video Streaming with Fixed Size of 10 MB.....	42
4.1.3 Short Video Streaming with Fixed Size of 50 MB.....	45
4.1.4 Short video streaming with variable size.....	46
4.2 Discussion .....	49
4.3 Summary .....	51
5. Conclusions and Future Works.....	52
5.1 Conclusion.....	52
5.2 Future Work .....	54
6. References .....	55

## **List of Tables**

Table 1-Number of nodes for each trace file .....	39
Table 2-Percentage of bandwidth usage out of effective bandwidth with fixed size of 10 MB file sizes.....	49
Table 3-Percentage of bandwidth usage out of effective bandwidth with Weibull distribution file sizes.....	49
Table 4-Percentage of bandwidth usage out of effective bandwidth with 50 MB files sizes .....	50

## List of Figures

Figure 1 - Six degrees of separation .....	9
Figure 2-Spotify Architecture .....	10
Figure 3 - Sources of data used by clients .....	11
Figure 4-The overlay network .....	12
Figure 5-P2P and client-server architectures .....	14
Figure 6-Hybrid P2P/Server suitable for streaming. In this figure, a variety of devices are included in the network. $\mu$ Torrent (BitTorrent Client) allows monitoring BitTorrent downloads via a cell phone. ....	18
Figure 7-The BitTorrent Architecture .....	20
Figure 8-Group Communication (Green circles denotes members of the same multicast group) .....	21
Figure 9-The Ring Topology .....	23
Figure 10-The fields appearing in a trace .....	24
Figure 11-Facebook model user network graph for 250 nodes .....	31
Figure 12-P2P networks of 250 nodes .....	32
Figure 13-Simulation Engine .....	34
Figure 14-Simulation Architecture .....	35
Figure 15-Experimenting on Trace1 with packet size of 512 bytes .....	40
Figure 16-Experimenting on Trace 2 with packet size of 512 bytes .....	40
Figure 17-Experimenting on Trace 3 with packet size of 512 bytes .....	41
Figure 18-Experimenting on Trace 4 with packet size of 512 bytes .....	41
Figure 19-Trace1-Short video streaming with constant size of 10MB .....	42
Figure 20-Trace2-Short video streaming with constant size of 10MB .....	43

Figure 21-Trace3-Short video streaming with constant size of 10MB .....	43
Figure 22-Trace4-Short video streaming with constant size of 10MB .....	44
Figure 23-Trace4-Short video streaming with constant size of 50MB .....	45
Figure 24-Trace 1-Short-video streaming with Weibull distribution of file sizes .....	46
Figure 25-Trace 2-Short-video streaming with Weibull distribution of file sizes .....	47
Figure 26-Trace3-Short video streaming with Weibull distribution of file sizes .....	47
Figure 27-Trace 4-Short-video streaming with Weibull distribution of file sizes .....	48
Figure 28-Percentage of bandwidth usage out of effective bandwidth with 50 MB .....	50

## Acronyms

<b>BT</b>	BitTorrent
<b>CBR</b>	Constant Bit Rate
<b>DDR</b>	Data Driven Randomized
<b>DHT</b>	Distributed Hash Table
<b>IM</b>	Instant Messaging
<b>IP</b>	Internet Protocol
<b>IPTV</b>	Internet Protocol Television
<b>ND</b>	Network Delay
<b>NL</b>	Network Latency
<b>Ns-2</b>	Network Simulator 2
<b>OSN</b>	Online Social Networking
<b>OTCL</b>	Object Oriented Text Command Language
<b>P2P</b>	Peer-to-Peer
<b>PDF</b>	Probability Density Function
<b>RFP</b>	Rarest first Policy
<b>RSS</b>	Really Simple Syndication
<b>TCL</b>	Text Command Language
<b>TCP</b>	Transmission Control Protocol
<b>TFTP</b>	Trivial File Transfer Protocol

**UDP**

User Datagram Protocol

**UGC**

User Generated Content

# Chapter 1

## 1. Introduction

People are connecting more and more with one another online. According to Gneiser *et al.* [1], “Web 2.0 is the use of lightweight, intuitive, Web-based services that rely on user participation and user-contributed data, and generally involve some level of social interaction and networking”. One of the most noticeable features of Web 2.0 is the introduction of online social networking (OSN). Social networking websites offer a highly dynamic and proactive setting for multimedia distribution.

### 1.1 Information Sharing in Web 2.0

Early generations of web sites were static, and interaction mainly consisted of email discussions and web-message forums. Communication on the internet was between web master and client, where web developers implemented websites that users visited and viewed.

Web 2.0 provides a new approach to information-sharing, which is based on the involvement of people and authorization of users. Social networking is an essential component of Web 2.0. As Alexander *et al.* mentioned in [2]: “Web 2.0 platforms are often structured to be organized around people rather than the traditional computer hierarchies of directory trees”. The success of a Web 2.0 page is reliant on how much users are contributing to it. Since many users contribute to the contents, there is a broad range of quality of information on Web 2.0 sites. Types of Web 2.0 components are as follows: [1]:

**Blogging:** users post materials such as text, images, audio and video files on a web site. Features of social networking can be seen in blogging.

**Wikis:** “A wiki is a database of pages which visitors can edit live” [3]. Wikis are websites where users contribute through a web browser by using a markup language.

**Tagging:** tags are an important feature of Web 2.0 services. A tag is a keyword or term assigned to a piece of information.

**Feeds:** RSS (really simple syndication) feeds allow users to see when content providers add new content to their web sites. RSS feeds can be in audio or video formats, or in the form of news headlines.

**Mash-ups:** a mash-up is a web application that consumes data or functionality from multiple sources in order to create a new service.

## 1.2 Social Networking Web Sites

With an overview of the Web 2.0 key components, subsection 1.2 will discuss how Facebook implements some of the Web 2.0 features. The ability to collect notes on Facebook, continuously update status and, express individual thoughts and opinions can be categorized as a *blog*. The equivalent of *Wikis* is the “groups” feature on Facebook where members can communicate with each other and exchange information with one another. Facebook users routinely upload *User-Generated Content (UGC)*. For example, users are continuously posting their material on Facebook including video, images and audio tracks. The “*News feed*” feature on Facebook behaves similar to RSS feeds. Considering different fan pages on Facebook, one could easily create a page and by simply liking the certain page, users can get information from the owner of the fan page. Such a feature can be also referred to as RSS.



### **1.3 Employing Social Networking in P2P Networks**

Since the 1990s, we have seen a vast growth of multimedia sharing and file sharing by P2P networks, starting with Napster [4]. Although Napster eventually failed, dozens of applications similar to Napster have since been launched. The rise of such applications has influenced the means through which multimedia files and information are exchanged. The creation of Napster-like applications has ultimately changed the music industry.

The first social networking web sites, starting with classmates.com, were established in 1995, but the early sites lacked many of the social aspects that have since been provided by Facebook. Unlike social-networking sites, the extent of social communication in P2P networks is the exchange of comments between users.

Spotify is the latest development of Sean Parker, the Napster creator and Facebook co-founder [5]. Spotify is similar to iTunes [6], with the main difference being that users can stream music via their computer or mobile phone. Spotify has a scalable P2P network that is able to serve the demand of users. Spotify has replaced the lack of social networks in P2P file-sharing applications by allowing Facebook and Twitter accounts to be integrated into Spotify. Users are able to log in with their Facebook accounts to share their music and playlists with their Facebook friends through their inbox.

### **1.4 Thesis Motivation**

Multimedia and information sharing in social networking has become increasingly popular; however, there has been little research that measures the performance of multimedia delivery in social networking web sites. Multimedia sharing in social networking sites such as Facebook falls within the client-server architecture.

In the study, we assume Facebook is integrated on top of a P2P network in such a way that each user is a peer in the P2P network, and can share information with his/her friend. Therefore, the Facebook user network graph is compared with other types of P2P networks.

Until now, there has been a lack of performance measurement of social networking when users employ a P2P structure in regards to multimedia and information sharing. In order to be meaningful, this performance measurement must be done by proper comparison of realistic scenarios. In the study of multimedia distribution, it is possible that the simulation result will be based on unfair comparisons. For instance, if a P2P network is compared with client-server architecture for the purpose of multimedia distribution, the consequent comparison scenario is an unfair scenario. Moreover comparing classical P2P networks and Facebook-based P2P network will provide necessary information for building an effective multimedia-sharing system.

Studies show that there is improvement in the performance of multimedia distribution when a P2P network is used [7, 8, and 9]. Considering the features and the benefits of P2P networks, it will be worthwhile to examine whether or not Facebook-based P2P video-streaming systems have the capability of providing a reliable and scalable platform for internet users to share any multimedia content.

There are very few simulators that can simulate the user networks in Facebook. The other motivation of this thesis was therefore to improve the Facebook simulator developed in the distributed system and multimedia processing (DSMP) lab at Ryerson University [10], to make it available as the open source to be used for research activities such as this thesis. In this thesis, a novel and fair comparison scenario is created, where users exchange information and share short videos with each other.

## 1.5 Thesis Objectives and Scope

P2P networks can be classified into three main categories: Decentralized-structured, Decentralized-unstructured and Centralized P2P. The objective of this thesis is the comparison of information-sharing scenarios by simulation of the Facebook Overlay Network, Structured P2P (Mesh Topology) and Unstructured P2P (Ring Topology) network. This objective is achieved through analysis of multimedia distribution and information transfer in the aforementioned networks.

In the scope of content delivery, this thesis focused on: 1) exchanging information; and 2) short-video sharing. “Exchanging information” refers to scenarios where users are writing comments and providing feedback to one another. As such, sets of experiments were designed where users exchanged information in the packet size of 512 bytes. Sharing short videos was accomplished by designing a P2P network where users exchanged: 1) a fixed-file size of 10 MB; 2) a variable-size file which was generated by Weibull distribution; and 3) a fixed file size of 50 MB. The 50-MB file size was chosen for video streaming in order to transfers large video files. Performing a set of experiments related to the short-video streaming in simulated networks shows how quickly information can be transferred in such networks.

There are several open source P2P simulators that can be used for implementing structured and unstructured P2P networks. A thorough discussion of P2P networks is presented in Chapter 2. After considering the network topologies that were to be simulated in this work, a suitable tool had to be chosen that included provisions for implementing the necessary features of these networks. Ns-2 was selected, as it considered all the network level details.

## **1.6 Research Contributions**

As previously mentioned, the objective of this work is to design and analyze information-sharing in a Facebook network that is based on P2P structure. The main contributions of this research are as follows:

- To the best of our knowledge, this work presents the first attempt to compare multimedia sharing in Facebook when using P2P networks;
- We found that the Facebook network that uses a P2P structure has the highest average throughput in the set of experiments performed for short video sharing. Therefore it is the most suitable network for sharing short video between users;
- Our achieved results in the set of experiments performed for exchanging information show that the P2P structure (ring topology) has the highest speed of information transfer compared to the other P2P networks;

## **1.7 Thesis Outline**

Chapter 2 will introduce the necessary background information regarding P2P networks and social networking sites such as Facebook and YouTube. In Chapter 3, we describe the functional aspects of our proposed model and simulations. Chapter 4 presents experimental evaluation and results of our model. Chapter 5 summarizes this work and provides suggestions for areas of future research.

# Chapter 2

## **2. Background Information and Related Works**

This chapter provides the necessary information about multimedia distribution in P2P networks, and discusses Facebook followed by Small World Phenomena, Spotify and YouTube. Furthermore, BitTorrent protocol and its components are explained in details. The chapter closes with the explanation of related work to this thesis.

### **2.1 Background Information**

#### ***2.1.1 Facebook***

Facebook is a social networking site where users interact through networks such as universities attended, friendships, interest groups, favorite movies, etc. Facebook contains more than 800 million users. Users need to register to use the web site, and then can add other people with whom they can then share messages, photos and videos. Originally, in order to join Facebook network users needed to be to be part of an institution, but currently anyone can join Facebook. Facebook web-based applications are growing significantly. One of the most interesting features of Facebook is that it now allows the integration of Skype video phone to its instant chat messaging service. Facebook emphasizes the fact that users do not have to create a new account to call each other, and this is made possible by the minimal setup for using Skype through Facebook.

Based on [11], applications on Facebook can be classified into the following categories:

- Friend comparison: applications that allows users to create list of top friends or best friends;
- Casual communication: users can write on each other's walls, send messages and in general communicate with each other;
- Rating, Taste Matching and Recommendations: users can recommend items to each other, ranging from restaurants to music.
- Gesture: Facebook contains an application that allows users to perform visual gestures.
- Self-Expression: users can express their moods and their opinions about different topics, etc.
- Gifting: users are able to exchange virtual gifts with each other.
- Meeting people: An application exists for people who are interested in online dating.

### ***2.1.2 Other Online Social Networks:***

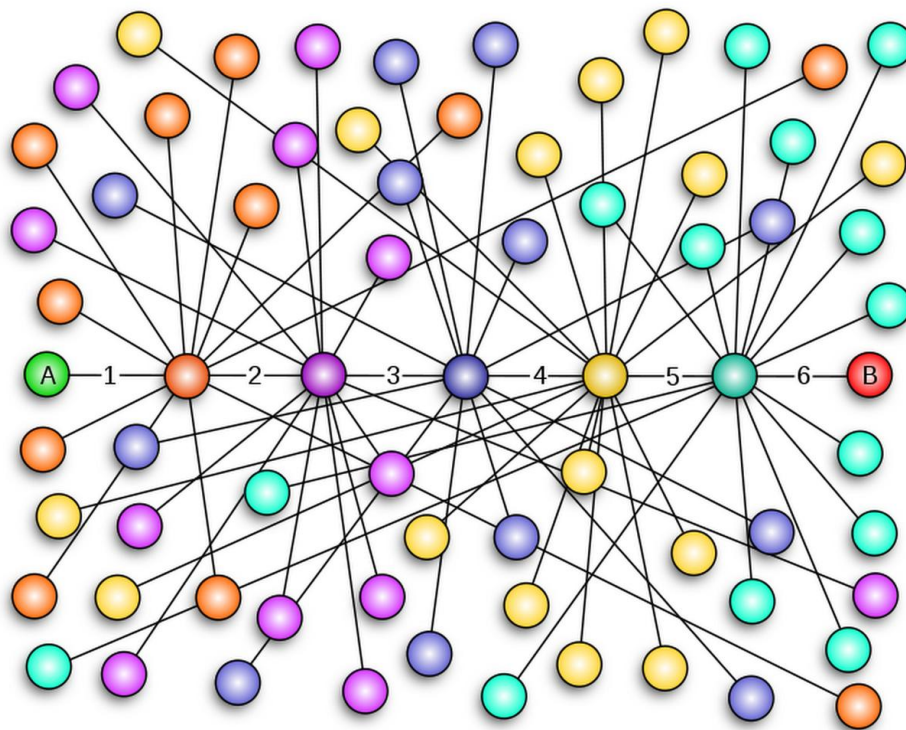
- Google plus [12] is a social networking web site created by Google. It has more than 90 million users.
- LinkedIn [13] is a business-related social networking site founded in 2002.
- Orkut [14] is a social networking web site operated by Google which is very popular in Brazil and India.
- Twitter [15] allows a user to send and read text-based posts of up to 140 characters, known as tweets.

### ***2.1.3 Small World Phenomena***

“Small world phenomenon” or “six degrees of separation” was introduced by Milgram [16].

Milgram's objective was finding a short chain of associates and linking the people who did not

know one another. “A social network exhibits the small-world phenomenon if any two individuals in the network are likely to be connected through a short sequence of intermediate acquaintances” [17]. Characterizing internet topological properties is necessary for the assessment of a new protocol, design choices, and large-scale content delivery mechanisms. The following figure describes the six degrees of separation:

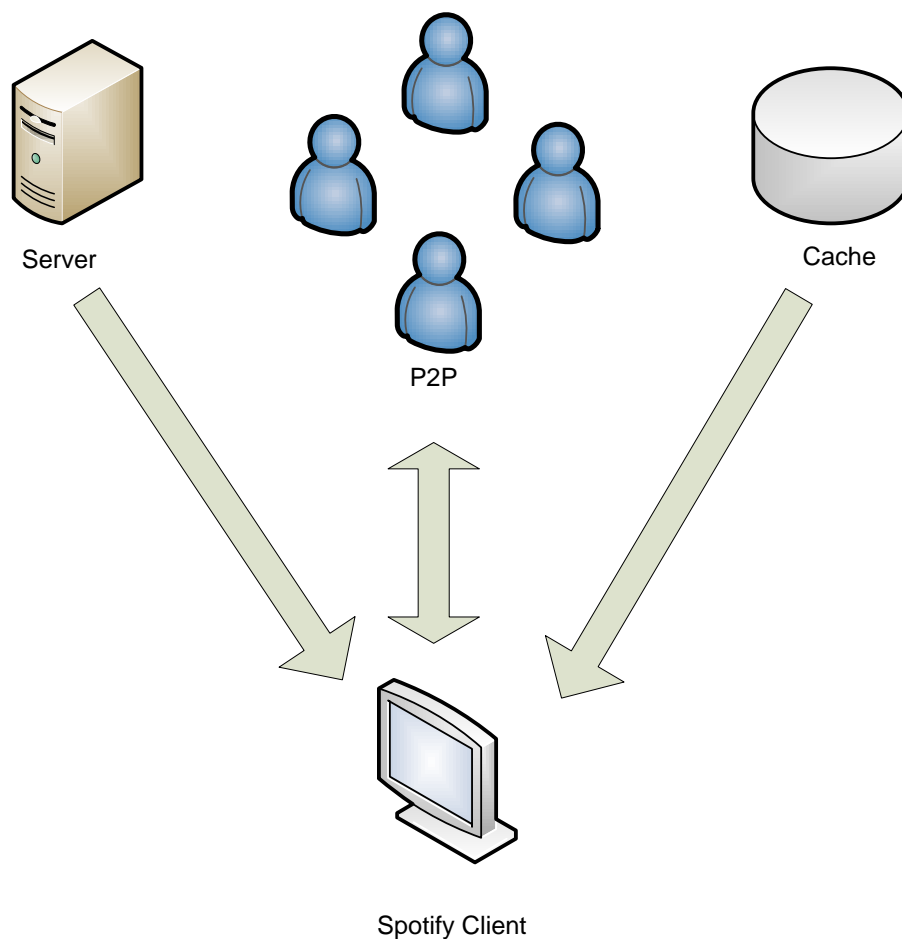


**Figure 1 - Six degrees of separation [18]**

#### ***2.1.4 Spotify***

Spotify is a music streaming service that offers access to a library of over ten million music tracks [19]. It is available in several European countries and it was launched in the USA in July 2011. A free version of Spotify is available that come with advertisement. There is also a

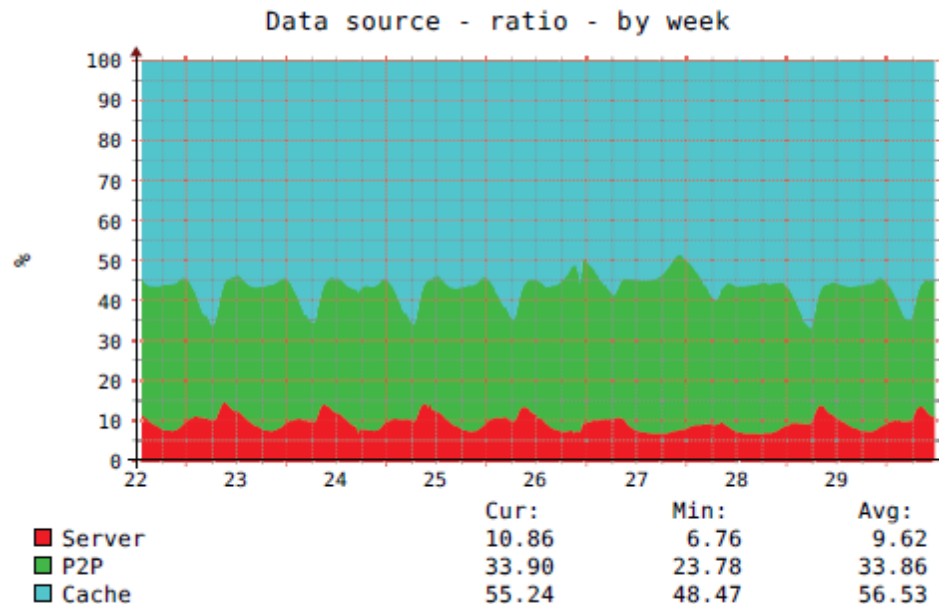
subscription service available. Users access Spotify with a client application which is available for Windows, Linux and OS X. Spotify is also available on several smartphone platforms, as well as hardware devices such as Logitech Squeezebox, Onkyo and Sonos [20]. Spotify protocol is a combination of client-server and P2P networks. The PC users form the P2P network. The Spotify scalable P2P network meets the requirements of users. The P2P network in Spotify offers low-latency streaming and moreover decreases the load on the Spotify servers. Users who are using Spotify with their smartphones are not participating in the P2P network, but only using the client-server architecture. Since the launch in the USA, social networking has been integrated into Spotify. All new users are required to access the service via Facebook login.



**Figure 2-Spotify Architecture**



With the integration of social networking, Spotify users and Facebook friends are able to share their playlist and tracks with each other. Tracks are transmitted in an encrypted format therefore users can only listen to the tracks in the Spotify. Users don't own the music unless they buy it.



**Figure 3 - Sources of data used by clients [19]**

According to [19], 8.8 % of the data are from servers, 35.8% are from the P2P network and the remaining 55.4% are cached data. Therefore the caches and P2P network together significantly decrease the load on the server.

### **2.1.5 YouTube**

YouTube was created in February 2005. YouTube allows users to share video files with each other. Features were added to YouTube in order to enable social networking between its users [21]. YouTube has received great attention from content providers, content consumers, and advertisers. The total views of YouTube are over 2 billion a day [22]. YouTube offers more than

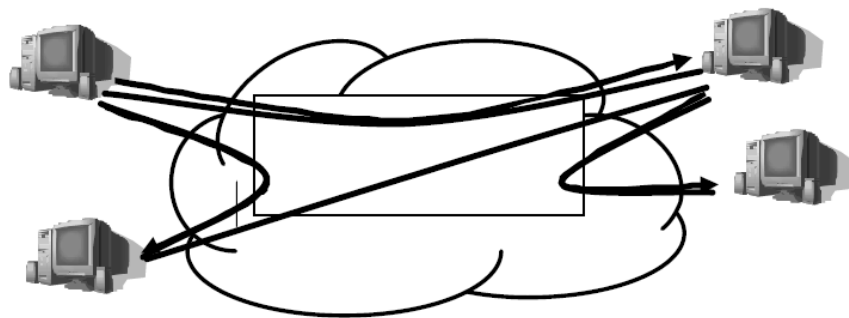
just video sharing. It has been used for exhibiting talent and building careers. In 2011, Google plus social networking web site was integrated with YouTube and the Chrome web browser, therefore making videos watchable within Google plus environment.

YouTube uses Adobe's Flash Video (FLV) format for video sharing. This is one of the reasons behind YouTube success. Users are uploading videos in different formats (MPEG, WMV and AVI), furthermore YouTube converts these formats into FLV before posting them into website Thus users are allowed to watch the videos without any additional plugins.

### ***2.1.6 Overlay Networks***

“An overlay network is a virtual network of nodes and logical links that is built on top of an existing network with the purpose to implement a network service that is not available in the existing network” [23].

Overlay networks are an old phenomenon. Internet originally was created as an overlay on top of the telephone network. The overlay network can control the sequence of nodes a message passes through until it reaches its final destination.



**Figure 4-The overlay network [23]**

### ***2.1.6.1 Two of the most popular approaches to constructing an overlay network:***

*Tree based:* Data is exchanged only on the virtual links of the overlay network; consequently using a distribution tree has a significant effect on the overall behavior of the overlay network. Nodes form a tree structure in order to transfer data, and parent-child relationships exist within the tree. Once a node receives data, it sends (pushes) a copy of that data to its children. Once a node fails, it must be repaired because its offspring cannot receive data anymore. If it is a parent node that fails, the consequences could be the loss of a vast amount of data [24].

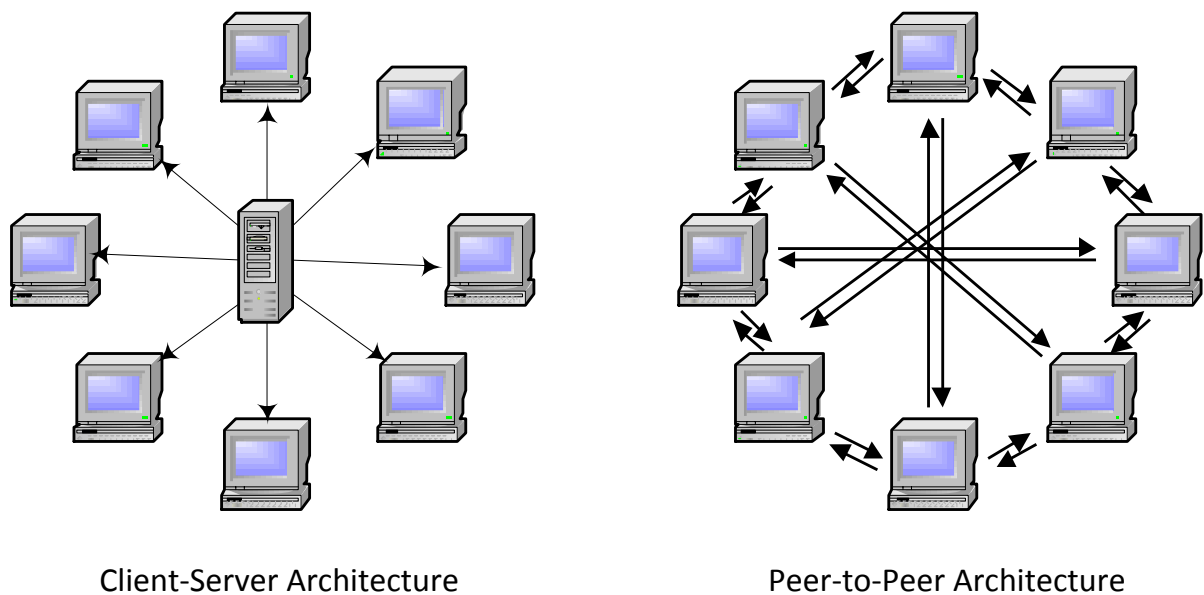
*Data-Driven Randomized (DDR):* In DDR there is no exact structure. The “gossip algorithm” is used, meaning a node sends the newly generated message to connected peers. Once other nodes receive the message, they transmit it to their peers and so on. This “random push” feature of the gossip algorithm can cause redundant transmission. To solve this problem, specific nodes create a group of partners. Therefore, a peer only shares information with one or more partners. In other words, a node doesn’t ask for data from partners unless it needs it [25].

### ***2.1.7 P2P Networks***

For more than a decade, P2P architecture has been more popular. P2P networks are distributed systems consisting of nodes in the network topologies with the determination of sharing resources. Content distribution, a notable application area of P2P systems, is built on systems designed for sharing multimedia files between users [26]. According to Shirky [27] “P2P is the class of applications that takes advantages of resources-storage, cycle’s content, human presence-available at the edges of the internet”. The most famous P2P file-sharing applications are gnutella network , eDonkey and LimeWire; however, the only application that relies on content replication is BitTorrent [28].

The aim of a P2P network is to create an environment where users can share large-sized files without creating a bottleneck. In P2P, structural responsibilities such as administration and maintenance are handled with users instead of a single unit. The key aspect is a suitable algorithm for the placement of information, receiving and sending information, and giving permission to the download resources located on computers throughout a network [28, 29].

P2P decreases the cost and bandwidth limitation of client-server architecture. Enormous numbers of users are entering and leaving the network. Any of these users could simply fail in terms of providing performance. Consequently the major difficulties for P2P architecture are scalability and self-organization without a server-based technology.



**Figure 5-P2P and client-server architectures**

P2P networks are divided into three categories: Decentralized-Structured, Decentralized-Unstructured and Centralized P2P [30]. This division is based on the connection of the nodes and object (file) query algorithms in overlay network:

- Centralized P2P: “Nodes in the P2P network issue queries to the central directory server to find which other nodes hold the desired files. Such centralized approaches do not scale well and have single points of failure” [31]. Napster is in this category.
- Decentralized-Structured P2P: nodes at the overlay network communicate in a deterministic procedure. A distributed hash table (DHT), which is a class of decentralized distributed system, is responsible for indexing and creating a random key for data items. What is challenging here is creating a good multi-paradigm programming language that uniquely maps the key of the data to the identifier of a node. DHT have also been used in unstructured P2P file-sharing applications such as Napster and Gnutella (Gnutella is an unstructured P2P network) DHT. Chord is an example of a structured P2P network.
- Decentralized-Unstructured P2P: the overlay network is shaped randomly. There is no algorithm defined for organizing the network connections, but rather it uses a flooding mechanism for searching a desired content. This way, when the peer submits the query it goes through the network to find as many peers as possible that have the content to share. If the data is popular, the query is most likely to become successful in finding a high number of provider peers. Examples of popular unstructured P2P networks are KaZaA, Gnutella [32], and BitTorrent.

Research shows that for today’s mass-market data-sharing applications, unstructured overlays perform better with more support than structured overlays, since peers are extremely transient [33].

### ***2.1.7.1 Chord***

As we mentioned above Chord is in the category of structured P2P networks. Nodes in the Chord network should know how to contact their current successor peer on the identifier circle. by assigning a key with each item and keeping the key/data item at the node data location is implemented easily. Chord can handle the dynamic network efficiently [34]. It can reply to queries once the nodes are joining and leaving the network. Furthermore Chord uses consistent hashing [35] to balance the load between the nodes. In the steady state while having  $N$  node in the network, each node keeps  $O(\log N)$  data about other nodes.

### ***2.1.7.2 Classification of P2P applications***

According to [26] there are four Categories of applications in the P2P networks:

*Communication and collaboration:* This class includes the applications that often provide real-time communication and collaboration between users' computers. Yahoo and MSN chat services are examples of this category.

*Distributed Computation:* This category describes systems that use other users' processing powers. A complex task of a peer is separated into different parts and distributed to several peers; peers will execute the task and return their results.

*Internet Service Support:* Based on P2P architecture, several applications have emerged that support internet services. For example, P2P multicast systems and applications providing protection against virus attacks.

*Database Systems:* Significant research has been done to create distributed database systems based on P2P architectures. Edutella is an open-source project which provides a metadata

structure and querying capability for P2P applications. PIER is a scalable distributed query engine made over a P2P overlay network which allows relational queries across multiple users.

*Content Distribution:* Most of the P2P file-sharing applications fall into this category, where users are sharing large and small files in different formats with each other. Media files, text files and applications could be among these files.

#### ***2.1.7.3 P2P Simulators***

PeerSim [36] is a P2P simulator written in Java. It has two simulation engines. One is a cycle-based engine which does not consider the details of the protocol stack. The other is an event-based engine, which is considerably slower but it allows for more accurate simulations.

GPS [37] is also written in Java. It is an event-driven, message-oriented simulator. GPS contains a graphical-user interface (GUI) in addition to the simulation components. The simulator is mainly used to simulate BitTorrent Protocol.

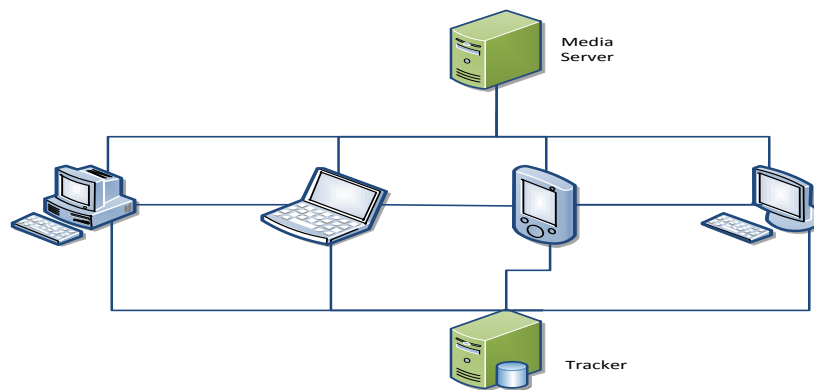
OverSim: “Oversim is a flexible Overlay Network Simulation Framework” [38]. OverSim [39] is based on OMNeT ++ and it provides simulation for both structured and unstructured P2P networks. Several overlay protocols are implemented in OverSim such as Chord, Pastry and GIA. These protocols can be used for both simulations and real-world networks.

#### ***2.1.7.4 BitTorrent***

Bram Cohen created one of the most successful P2P applications called BitTorrent (BT) [40]. The idea was to create an application which could easily update and download large-size files. It is apparent from recent literature that BT has become one of the most successful P2P applications, primarily due to its high efficiency and outstanding scalability [41].

In previous applications of P2P, the shared files were small, typically MP3 files, but BT can transfer huge files such as movies and TV series.

**Tit-for-Tat schema:** Preventing freeriders (clients who act selfishly and only download without uploading) has been a key concern with P2P applications. In previous P2P applications, a user was able to download at a very high rate depending on the bandwidth. However, the upload could be slower. BT Tit-for-Tat policy makes the download rate proportional to the upload rate [42, 43]. In this way fairness is achieved and clients who are downloading the same file at the same time are profiting from each other (Tit-for-Tat has had a significant impact on popularity of BT). Figure 6 describes a hybrid P2P/Server:



**Figure 6-Hybrid P2P/Server suitable for streaming. In this figure, a variety of devices are included in the network.  $\mu$ Torrent (BitTorrent Client) allows monitoring BitTorrent downloads via a cell phone.**

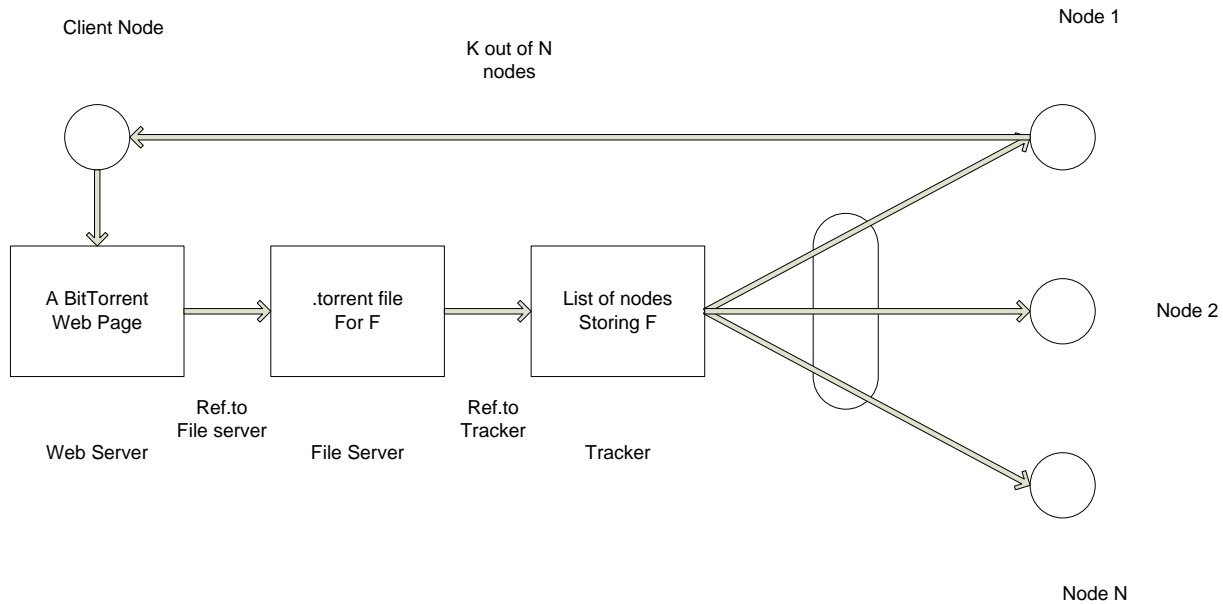
**Tracker:** Tracker is the only centralized component of BT protocol. It returns a list of peers. This helps the peers to find each other, and to keep the upload/download statistics of each peer. Tracker provides IP addresses for peers in the network [44]. Tracker does not involve itself in the content distribution.



**Rarest-First Policy:** BT gives higher priority to download the pieces which are rarest in the network. In such a way the uncommon pieces are gathered first. This method is called “rarest first.” This system also makes sure that pieces that are more popular are left for later. Assuming number of peers including segments is  $P(I)$ , if segment  $m$ ,  $p(m)$  is less than  $p(n)$  and  $n$  is not equal to  $m$ , the segment  $m$  is downloaded first. In this way receiving the entire file is guaranteed [42].

**BT Client:** BT client is any application that implements the BT protocol. Each client is capable of requesting and transmitting any type of computer file over the network using the protocol. Popular BT clients are Bitcomet (implemented with C++) and Vuze (implemented with Java).

**The BT Network or Swarm:** The BT network, also referred to as a swarm, does not provide any resource query, look-up, routing or topology-forming functionalities. Swarm is the group of peers that would include leechers, seeders and trackers. The main purpose of the swarm is to disseminate data. As more peers join the swarm the likelihood of successful download increases. The system is able to distribute large pieces of data such as compressed video files; interested peers connect to several other peers simultaneously and download different pieces of the file from other peers. Figure 7 shows the general architecture of BitTorrent:



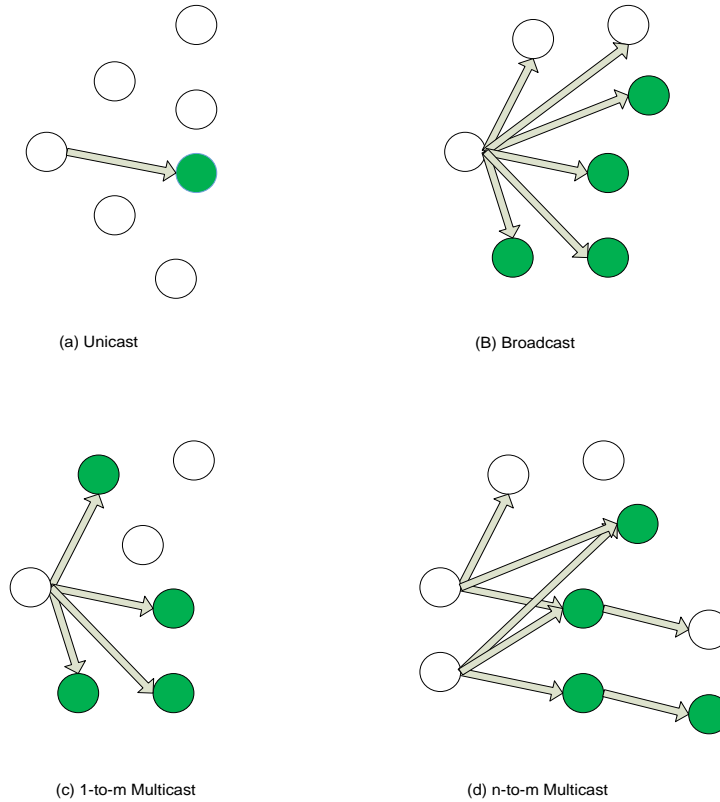
**Figure 7-The BitTorrent Architecture [29]**

#### **2.1.7.5 Distributed Hash Tables**

DHT is a method that addresses the problem of scalability and the efficient location of data items. DHT belongs to the category of decentralized distributed systems. It provides a look-up service similar to a hash table where pairs (key, value) are stored in a DHT. A DHT network performs three basic operations associated with <key, value> pairs: insertion, look-up and detection. DHT delivers scalable routing and indexing, robustness, and low-latency properties [45, 46].

### 2.1.7.6 Group Communication

IP multicast (IPMC) is usually deployed in multimedia content delivery networks. Here is a brief discussion about the technical aspects of group communication as implemented by IPMC [47, 48].



**Figure 8-Group Communication (Green circles denotes members of the same multicast group) [48]**

**Unicast:** The destination address is a single host. Communication is between a single sender and a single receiver.

**Broadcast:** The broadcast is directed to all hosts on the network and all the host must participate in the broadcast.

**Anycast:** The communication between a single sender and the closest of several receivers in a group.

**Multicast:** delivers an effective way of distributing data from a source to all the members in a multicast group. In multicast the source send a single copy to all the members.

*Here are lists of relationships between Multicast groups:*

**One-to-Many:** one host is sending data to many recipients, creating the multicast group.

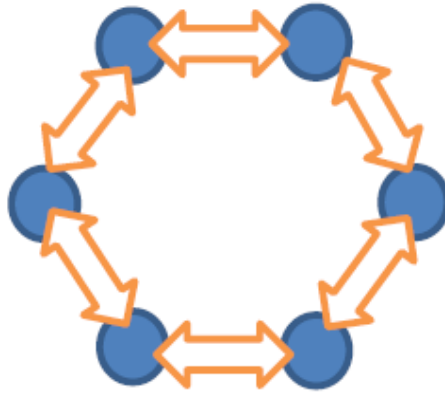
**Many-to-Many:** several sources are sending to the multicast group.

**Many-to-One:** Many to one is not a multicast relationship but an arrangement to distinguish between applications. It can be viewed as the response path sent from a One-to-Many Multicast network.

### ***2.1.8 Ring***

Ring is a network structure where network nodes are connected to each other, forming a large circular shape. The “ring topology” is considered cost-effective sustainable network architecture due to bandwidth sharing and increased survivability. Each packet is transmitted around the ring until it reaches its final destination.

Ring network or ring topology could be offered as a solution for group communication. Ring has major advantages over mesh and tree-overlays in terms of reliability and security [49]. Traditional tree and mesh-based networks have inherent flow and congestion-control problems [50]. Ring topology has certain advantages in terms of reliability and single-fault tolerance because of the structure of the network, where packets are looped back to the sender. Ring network simulated in this work, consists of bidirectional connections. Here is an example of a ring topology:



**Figure 9-The Ring Topologyi**

The ring topology does not require a central node to organize the connection of its users. This is one of the advantages of a network with a ring topology. Conversely, if one node fails, it creates a problem for the entire network.

### ***2.1.9 Network Simulator 2(Ns-2)***

Ns-2 is an open-source simulator that supports a wide variety of network protocols. Ns-2 is based on two languages, C++, and object-oriented text command language (OTCL). Data collection in Ns-2 is supported by traces. Traces keep track of the events related to the generation, enqueueing, forwarding, and dropping of packets. Trace files consist of lines of ASCII characters that have information related to the packets. Packet size, source/destination addresses and the types of protocols are contained in the trace files [51, 52].

Figure 10 describes the information contained in the trace files.

Event	Time	From node	To node	Pkt type	Pkt size	Flags	Fid	Src addr	Dst addr	Seq num	Pkt id
-------	------	--------------	------------	-------------	-------------	-------	-----	-------------	-------------	------------	-----------

**Figure 10-The fields appearing in a trace [52]**

**Network Latency:** (NL) is the time it takes for a message to be transferred from sender to the router, and also the time it takes to return that message. Network latency can be used as a measure of how fast a network is running. “Ping” is usually used to calculate round-trip latency. NL at network level can be caused by bottlenecking or overextension. One of the best ways to identify the nature of the problem is by trace routing. Trace route can analyze the movement of packet transfer along a network, so that the traffic can be re-routed while the badly-behaved area is rearranged [47]. There are several contributions to network latency. Here is the list of them [47, 48, and 53]:

**Queuing delay:** Once a router receives packets, they need to be processed and transmitted. If packets are received faster than the router can process them, they are buffered. Queuing delay is proportional to buffer size. The greater the numbers of packets to be transmitted, the longer the average wait (the time it takes for packets to be executed in router’s queue).

**Processing delay:** The time that a router needs to process the packet header (packet header is the data placed at the beginning of the block to be transmitted). Here the router checks the packets for any bit-level errors.

**Network Delay:** The time it takes for a bit of data to transfer from one node to another. In general, users care about the entire delay of the network, which is overlay delay (OD). Network delay is considered as a subset of overlay delay.

**Transmission delay:** The time it takes to push all the data into a link:  $D_T = N/R$ . Where  $D_T$  is the transmission delay,  $N$  is the number of bits and  $R$  is the rate of transmission.

## 2.2 Related Works

In this section related works about social networking and P2P networks are explained.

### 2.2.1 Hybrid P2P Systems

Bashardoust *et al.* [54] considers the effect of the content delivery network architecture on the popular short-video sharing website, YouTube. Due to increase in number of users in YouTube, scalability becomes a matter of importance. Therefore [54] claims that YouTube scalability requires a new content-delivery structure. Furthermore they propose a P2P network for short-video sharing where all participants' nodes are responsible to distribute video replicas they have stored. The system is a BitTorrent-like network in which a central media server is added. The proposed model improves the download bandwidth of client peers and load on the server, thus showing continuous playback without disruption for users.

### 2.2.2 Multimedia Characterization in Social Networking Web 2.0 Sites

There have been numerous studies relating to multimedia characterization in traditional web contexts. A comprehensive study on multimedia-streaming characteristics was done by [55]. Some of these research studies, such as [56] and [57], analyzed traces of streaming media and used the results to develop multimedia-workload generators. There are only a few research studies related the analysis and characterization of Web 2.0 sites, which is our area of focus. A

YouTube traffic characterization is presented by Gill *et al.* in [21]. Huang *et al.* [58] analyzed the potential benefits of peer-assisted video-on-demand service using the nine-month MSN video trace. Halvey *et al.* [59] provided an analysis of the social interactions on YouTube to understand community behavior. Mislove *et al.* [60] studied four online social networking sites, including Flickr, YouTube, LiveJournal, and Orkut. Also Paolillo [61] investigated the social structure of YouTube, addressing friend relations and their correlation with tags assigned to uploaded videos. Zink *et al.* [62] analyzed the content distribution in YouTube, and then performed a measurement study of YouTube traffic in a campus network. Cha *et al.* [63] considered file referencing behavior of user generated content in more detail. They sampled Daum UCC [64], the most popular UGC service in Korea, and YouTube repository, by using web crawler to study file referencing patterns. The study of Cheng *et al.* is reported in [65]. They collected information of YouTube videos in a three-month period to obtain characteristics of YouTube workloads. The authors concluded that utilizing a delivery method based on P2P architecture for YouTube could be challenging and make the situation even worse. However, their social network findings drove them to propose a novel P2P system based on the social network for YouTube.

### ***2.2.3 Analyzing YouTube and Facebook***

Workload characterization by Soraya *et al.* [66] studies the most popular short-video sharing service of Web 2.0, YouTube. Characteristics of 250,000 popular and regular videos were analyzed during five months. The main impact of [66] is presenting the essential elements required for development of a realistic workload generator for a short-video sharing site. Her work proves that caching the most popular videos in the proxy and prefetching the related videos in the client site can decrease the start-up delay.



Pakzad *et al.* [10] has modeled the user networks and social characteristics of Facebook. We have used the synthetic workload generator of [10] in our simulation. Based on the data, gathered during seven months of research at the DSMP lab at Ryerson University, Pakzad *et al.* [10] found that the user network node degree distribution in Facebook follows a lognormal distribution, and the distribution in which the number of users' fan pages increases, follows a Weibull distribution. Eventually software was implemented that creates graphs with similar static and dynamic features as users' networks on Facebook.

#### ***2.2.4 Multimedia sharing in overlay networks***

Although sharing information in the form of text and images has been very popular on Web2.0 social networking sites, only recently has video sharing started on Facebook. However, all types of multimedia sharing based on P2P structure have been proposed and used widely in recent years. The P2P files sharing applications such as BitTorrent, Skype, Napster, PPLive and Gnutella can be used for multimedia and gaming as well. There are several P2P system proposals to provide video on demand, such as [67, 68, and 69], but there are not as many proposals as there are for live streaming.

# Chapter 3

## 3. Proposed Model and Methodology

In this chapter we are going to describe a model for online social networking that can share multimedia and information similar to P2P networks. Note that the important social networking web sites such as Facebook and YouTube are based on client-server architecture; we focus on Facebook to describe our model. Users will be able to share multimedia in the application layer through the user's networks in Facebook when those networks are placed in the application level and on top of UDP. (The transmission protocol that we proposed to be UDP is discussed in the next session).

### 3.1 Integration of Social Networking with P2P Network

Spotify uses a Facebook users' network (by requiring users to integrate their Facebook account) on a P2P network on top of TCP. One of the ways of using a P2P network with Facebook is to integrate users' Facebook accounts on an available P2P network. This way, the users are able to have their Facebook friendship network and share multimedia with their friends. This resolves the first step of a P2P network's life cycle since the overlay network is formed by this approach, and a user has access to his/her Facebook friends through such a network. Spotify has recently used this type of integration. Spotify uses TCP because it is designed for music sharing that may not include video, so it can take the advantage of TCP reliability. However our proposed model is different from Spotify because for information transfer we suggest using UDP instead of TCP. We propose UDP to be used for information transfer in our proposed model because it is faster and more suitable for video sharing.

“UDP is defined to make available a datagram mode of packet-switched computer communication in the environment of an interconnected set of computer networks” [70]. In TCP, sending acknowledgment will delay processing the files. We chose to use the UDP instead since it comes with constant bitrate (CBR). UDP can therefore be used for streaming media applications, online games, IPTV, Trivial File Transfer Protocol (TFTP) and voice over IP. UDP sends the media stream as a series of small packets. We used UDP in all three network topologies.

Considering nodes’ lifetime cycle, there are four phases for P2P a network: join, query, download and depart [71]. In the first phase, a node joins a P2P network. So a node can get some basic information for startup, as is done in Facebook: the node can use its friends for sharing information. In the second phase, a node requires watching a video or sharing information. Here the Facebook user network is used to create the source and destination nodes. Third, the nodes will start to exchange files with each other. In this thesis, the P2P simulation emphasizes on the third phase, which is the download phase.

## **3.2 Simulating Facebook Network**

The Facebook network required for our simulation was generated based on the Facebook simulator explained in [10]. In this work all data and node connections were gathered over seven months in the Distributed System and Multimedia Processing (DSMP) lab at Ryerson University. The results of [10] shows that the node degree distribution of Facebook fits the lognormal distribution.

The data set consists of numbers of users’ friends and lists of their friends. This data set is made of real data from the Facebook network. To increase the reliability of our data set, three artificial data sets were also generated using the Facebook simulator.

For generating multimedia sharing, we used Weibull distribution for generating proper video file sizes between nodes. This course of action was based on a previous study [66] that found Weibull to be the suitable distribution to generate random video file sizes used in YouTube.

The Facebook simulator uses lognormal distribution [10]. “A random variable  $X$  is log-normally distributed if  $\log (X)$  has a normal distribution” [72]. The lognormal distribution is shown by  $\ln N (\mu, \sigma^2)$ , where  $\mu$  is a location and  $\sigma$  is scale. The PDF of the lognormal distribution is calculated by the following formula:

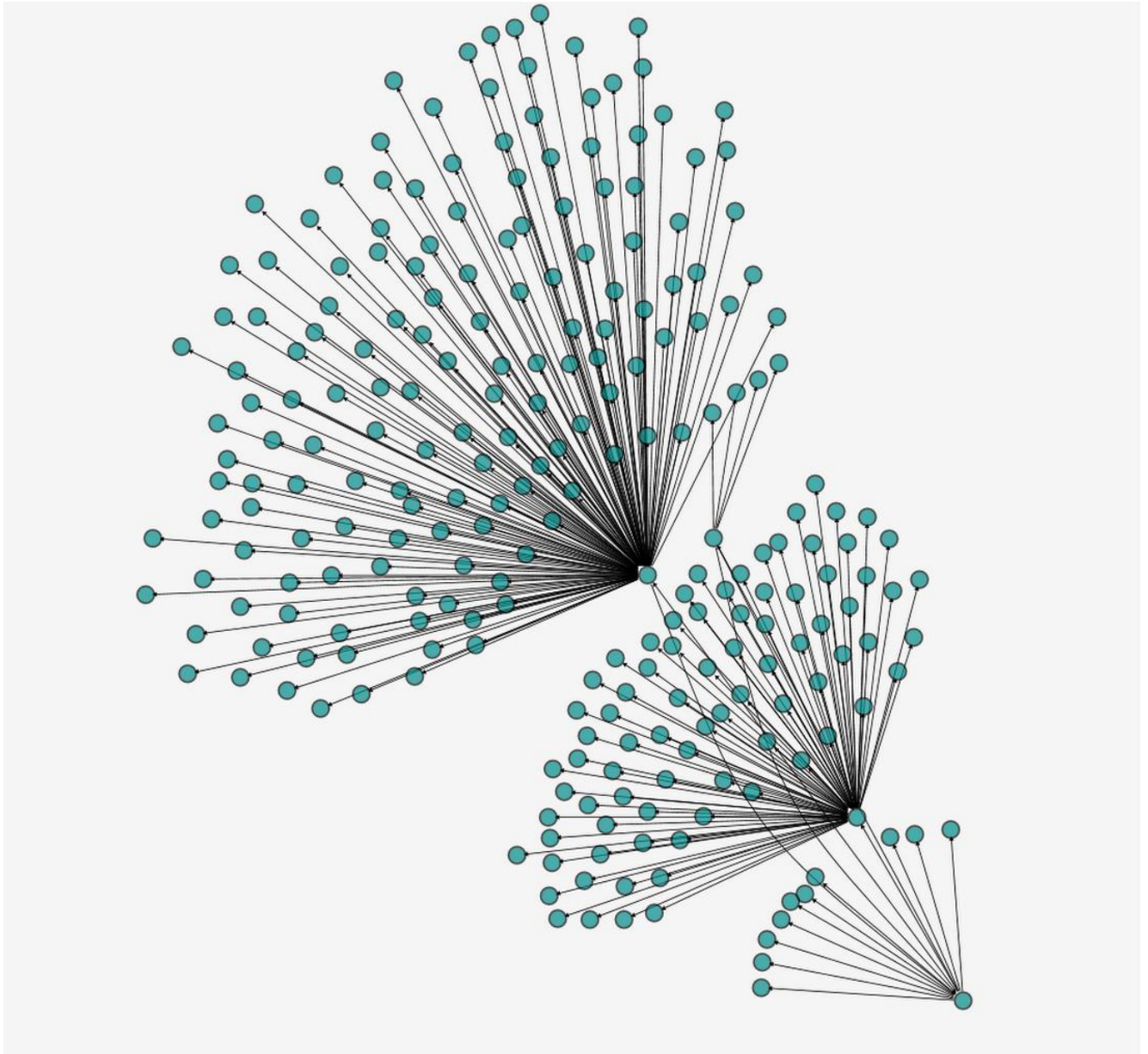
$$F_x(x; \mu; \sigma) = \frac{1}{\sigma x \sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} \quad x > 0 \quad (3.1)$$

For this work the lognormal distribution that is used by the simulator has  $\mu$  equal to 5.41 and  $\sigma$  equal to 1.19.

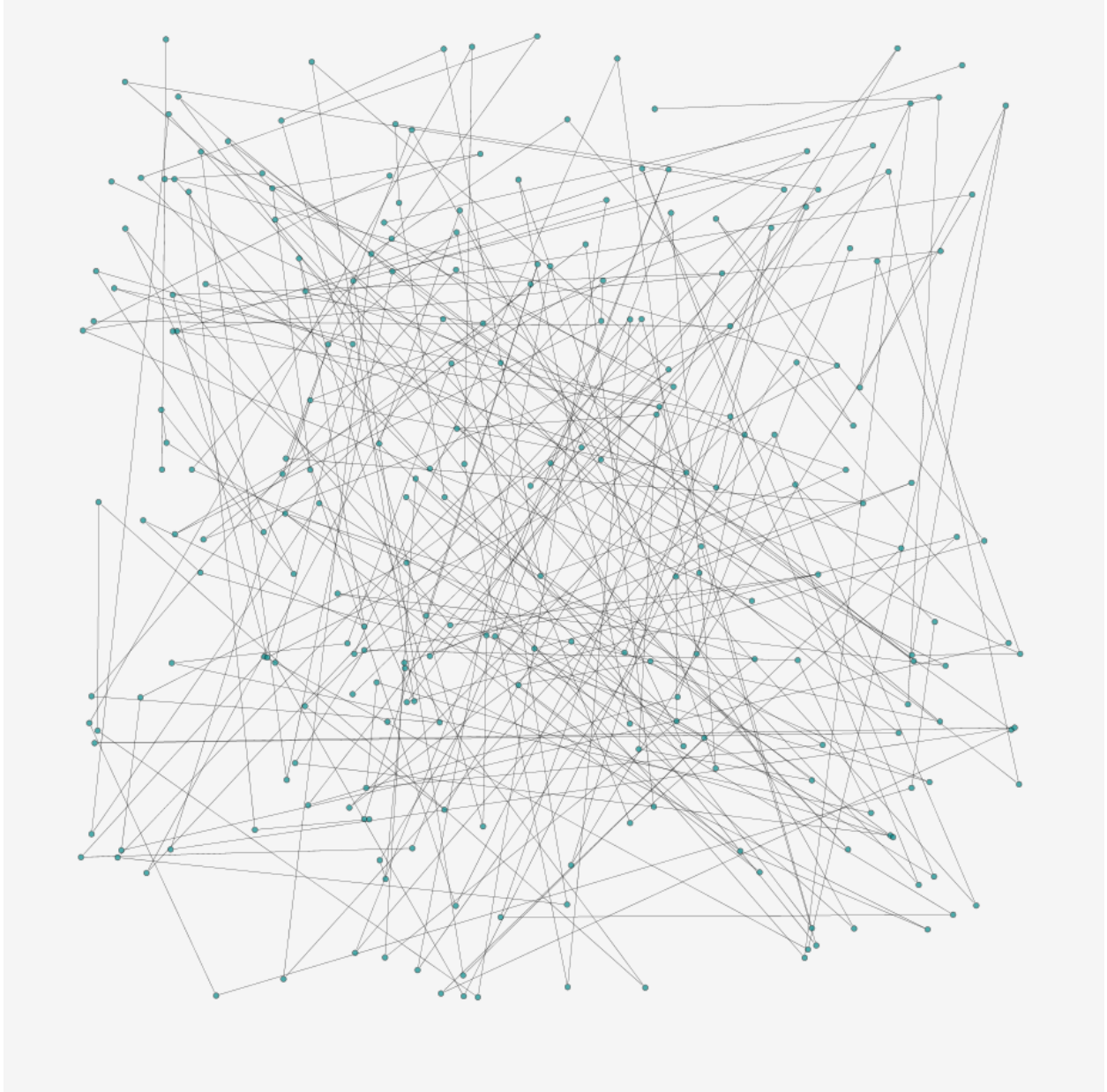
In the first data set, which is from real data gathered by the Facebook simulator, the Facebook simulation contains 1031 nodes in the network. The 1031 nodes are integrated in the simulation as a first trace file. The second and third trace files consist of 975 artificial network nodes and the fourth trace file consists of 500 nodes, made by the Facebook simulator. In our work, we measured the average throughput of each network along with the information transfer to assess the performance of each of the three topologies.

Users need to login to the Facebook web site in order to interact with each other. We are assuming all the users in the simulated Facebook are logged in and online. Information transfer in Facebook is similar to chat services such as MSN or Yahoo messengers. The following figure

shows the 250 Facebook nodes from the real data that is integrated in our simulation. Figure 11 and 12 are created by cosbi visualization tool [75]:



**Figure 11-Facebook model user network graph for 250 nodes**



**Figure 12-P2P networks of 250 nodes**

Ns-2 simulates random traffic for P2P networks. The random traffic distribution that is used in Ns-2 for connecting the nodes was not suitable for simulating over P2P networks. Therefore the program was modified.

In our simulation, connections of the nodes are based on uniform distribution for generating unstructured P2P networks. In the modified program, timing for sending packets is

also random, generated by uniform distribution as well for all three topologies. Therefore, nodes are sending packets at random times and this is applied for Facebook, unstructured P2P and structured P2P. Facebook, P2P structured and P2P unstructured all have bidirectional connections.

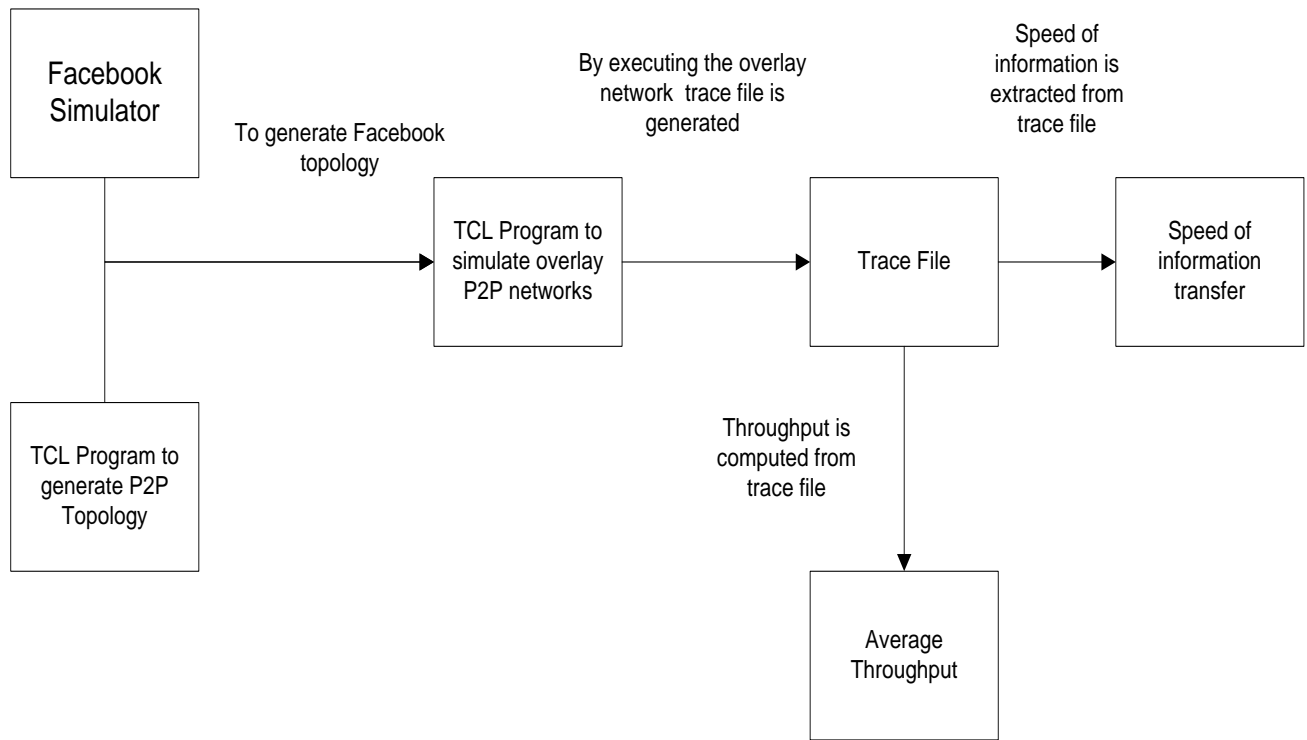
### **3.3 Simulating P2P Networks**

For structured P2P we use a ring-based topology and for unstructured P2P a mesh-based topology. The assumption is searching and routing of messages are done and the download phase of P2P system is in effect. In the ring topology each node is connected to its successor and its predecessor.

#### ***3.4.1 Simulation Architecture***

The simulator works in the following way: Upon the execution, the main program asks for the number of nodes. Once the overlay P2P networks are created, the Ns-2 link agent creates traffic between the nodes. As mentioned before, the connection between the nodes in the Facebook network is predefined by the Facebook simulator [10]. Ns-2 executes the P2P network and the outputs are four trace files. From the trace files speed of information transfer and average throughput are computed. Furthermore in our experiments we will use the YouTube simulator to distribute YouTube-like file sizes in our network, which will be discussed in detail in Chapter 4.

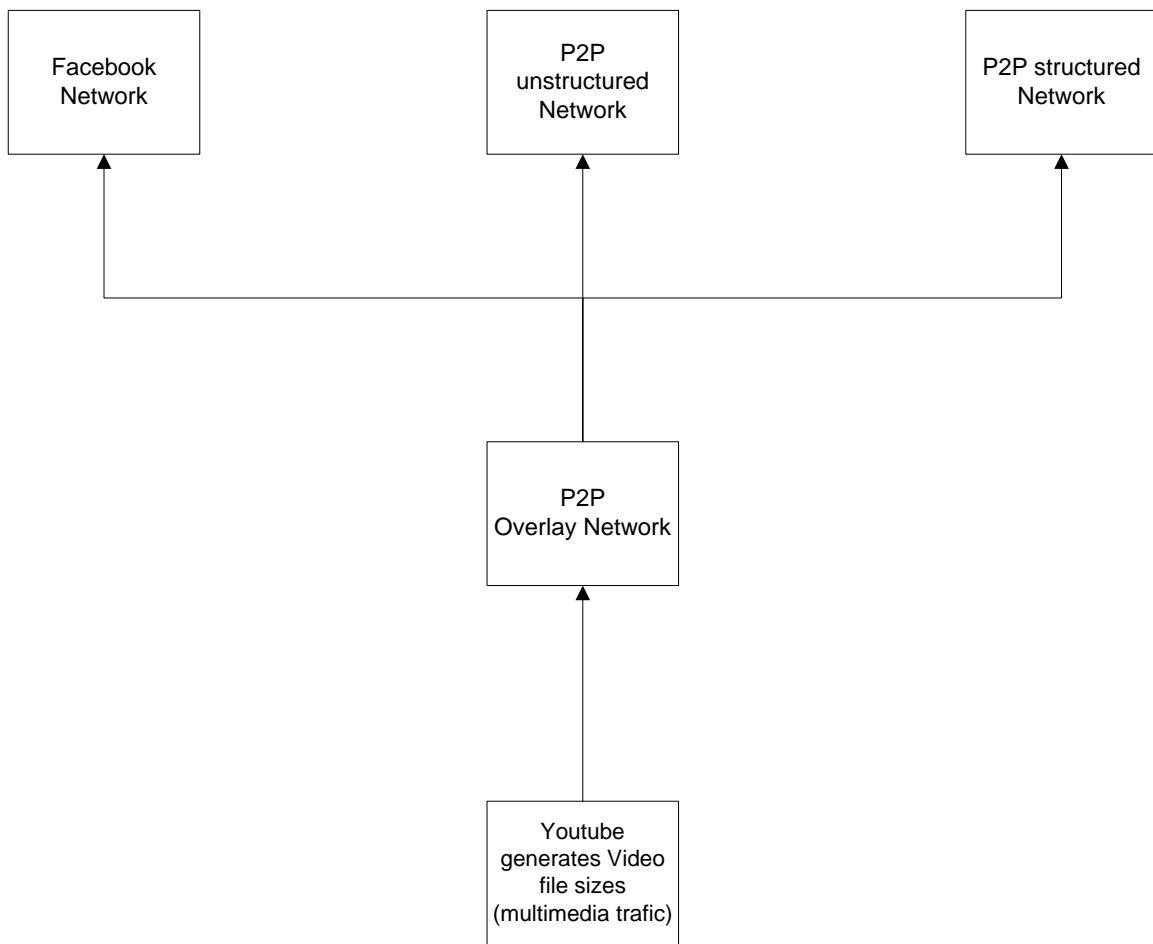
Figure 13 and 14 describes the simulation engine and simulation architecture:



**Figure 13-Simulation Engine**



The overlay networks are the P2P-based communication networks which will be used for multimedia and information testing performed by Ns-2.



**Figure 14-Simulation Architecture**

The Weibull distribution is a continuous probability distribution. The probability density function (PDF) of the Weibull distribution is calculated with the following formula:

$$f(x; a; b) = \begin{cases} ab^{-a} x^{a-1} e^{-\left(\frac{x}{b}\right)^a} & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (3.2)$$

Where  $a$  is the shape and  $b$  is the scale parameter of the distribution.

For generating multimedia traffic Weibull distribution is used with shape parameter  $a = 1.13$  and the scale parameter  $b = 11,193$  based on [66].

### 3.4 Performance Metrics

Finally for comparison between overlay P2P networks the average throughput is used. Throughput is the quantity of how fast files can pass through an entity (such as a point or network). Moreover, if we study this entity as a wall through which bits pass, throughput is the number of bits that can pass this wall in one second.

Several attempts were made for calculating the average throughput and speed of information transfer, such as creating the throughput for a specific node in different time periods. Finally average throughputs of the entire network along with the speed of information transfer are computed.

Formula 3.3 and 3.4 describes these performance metrics:

$$Throughput_{Avg} = \frac{\alpha}{\tau} \quad KB/s \quad (3.3)$$

$\alpha$  = traffic between nodes(KB)

$\tau$  = Simulation Time(Seconds)

$$Speed\ of\ information\ Transfer = \frac{\gamma}{\tau} \quad Nodes/s \quad (3.4)$$

$\gamma$  = Total number of unique nodes(nodes)

### 3.5 Summary

The architecture of our model for the Facebook and P2P network topologies are explained here. Average throughput of the network and the information transfer are mentioned in this chapter as the performance metrics. It is clear that the higher the throughput, the better the network is suited for multimedia file transfer. With Ns-2, these topologies are simulated as the download phase of P2P networks. Based on these models, the simulation scenarios for social networking and P2P networks are implemented and compared by other P2P networks.

# Chapter 4

## 4. Experiments and Results

In this chapter, all the experiments are presented, followed by the results. In order to complete these set of experiments three machines were used: firstly, Ryerson Dyno server which runs Ns-2 simulator; secondly, an Intel dual core system with 4 GB of RAM; and thirdly, an AMD six-core system with 16 GB of RAM. Our results in this chapter are also published in [73]. (The title of the paper is “Which type of Overlay Network Performs better for Multimedia Distribution and Information Sharing?”)

In this chapter unstructured P2P is referred to as P2P and structured P2P is referred to as ring network.

### 4.1 Creating Facebook, Peer-to-Peer and Ring Overlay Networks

For simulating an unstructured P2P network, Ns-2 was configured to generate a network structure with random topology. For simulating ring network, Ns-2 was configured to connect each node to the next node sequentially, meaning that each node is connected to its successor node and its predecessor node. We used Ns-2 to generate random P2P and ring networks; however, for simulating Facebook network we were not able to use Ns-2 on its own, so Ns-2 along with the Facebook network generator were used in two different approaches. Firstly, the real data gathered at the DSMP lab at Ryerson was used in Trace 1. Secondly, a Facebook simulator [10] was used for generating Trace 2, Trace 3 and Trace 4. Trace 2 and Trace 3 results look very similar since both trace file have the same number of nodes however they are two different data sets.

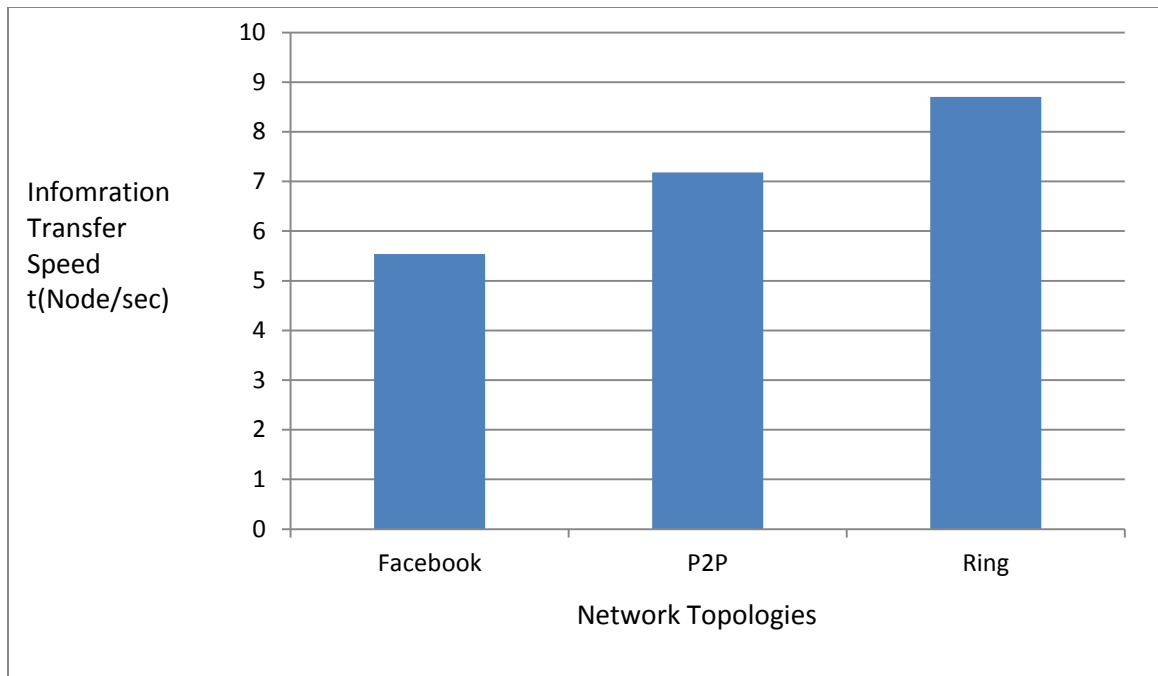
**Table 1-Number of nodes for each Data set**

	Number of nodes in each trace file for all topologies
Data set 1	1031
Data set 2	975
Data set 3	975
Data set 4	500

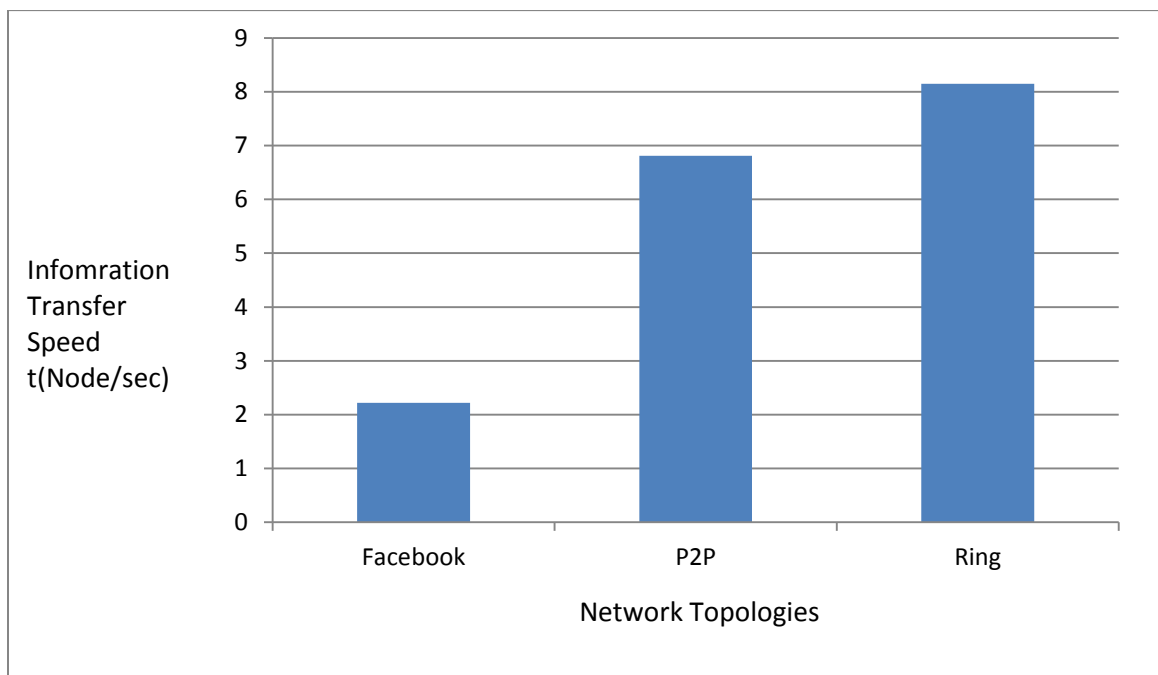
#### ***4.1.1 Simulation of Simple Information Transfer***

Experiments begin with simple information transfer where nodes are exchanging packet size of 512 bytes. The packet size of 512 bytes is chosen to represent the exchange of feedback by users, such as comments and other posts for sharing their opinions. Speed of distribution is the performance metric used in this specific experiment. The results of this experiment are shown in the Figure15, Figure 16, Figure 17, Figure 18 and Figure 19.

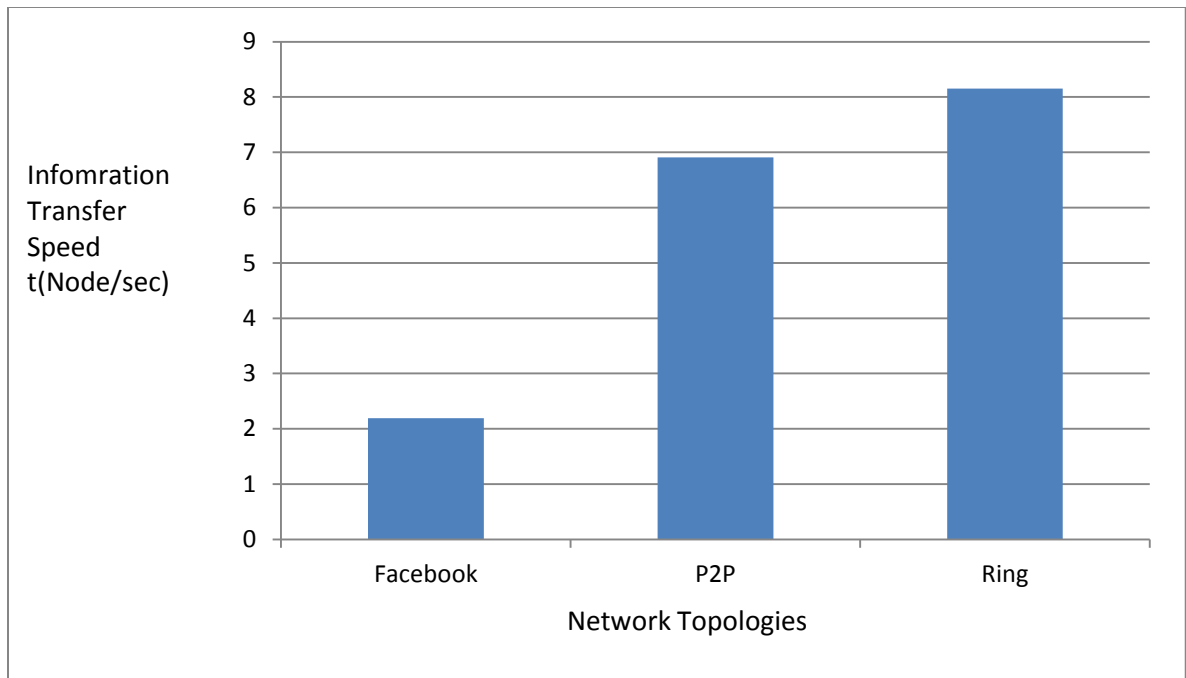
The details about the Ns-2 trace file are contained in Chapter 2. For all the experiments we are considering bandwidth of 2MB and Queuing Delay of 2ms. The first three experiments indicate that the Unstructured P2P (ring topology) is the fastest in terms of distributing information between users.



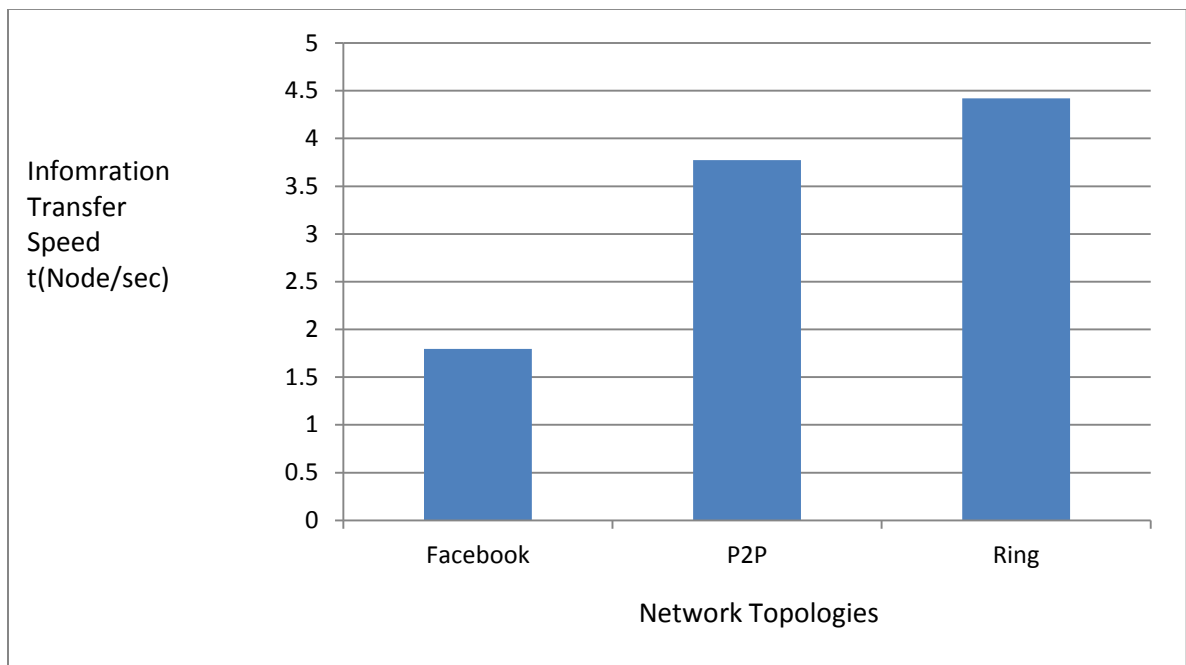
**Figure 15-Experimenting on Trace1 with packet size of 512 bytes**



**Figure 16-Experimenting on Trace 2 with packet size of 512 bytes**



**Figure 17-Experimenting on Trace 3 with packet size of 512 bytes**



**Figure 18-Experimenting on Trace 4 with packet size of 512 bytes**

#### 4.1.2 Short Video Streaming with Fixed Size of 10 MB

In our second experiment, nodes exchanging a fixed-size, short video of 10 MB. The next three graphs show the video sharing based on fixed size file of 10 MB for each trace file, where Facebook has the highest average throughput. By setting the file size to Weibull distribution for each trace file, a similar trend is observed, and Facebook continues to have the highest average throughput.

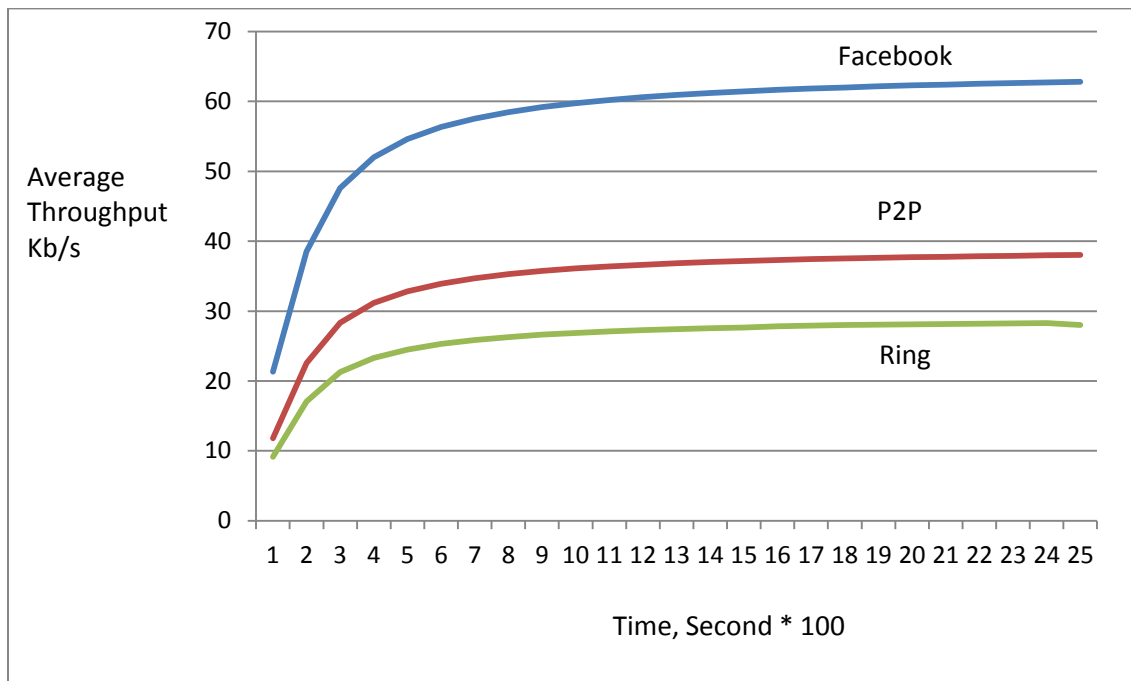
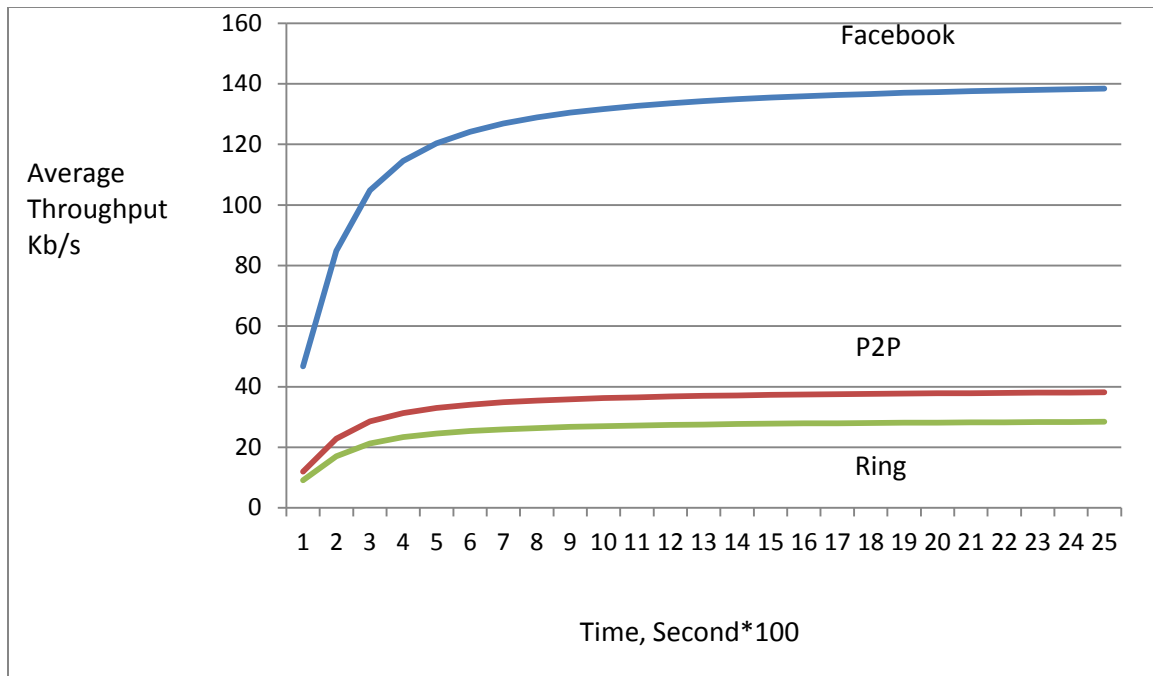
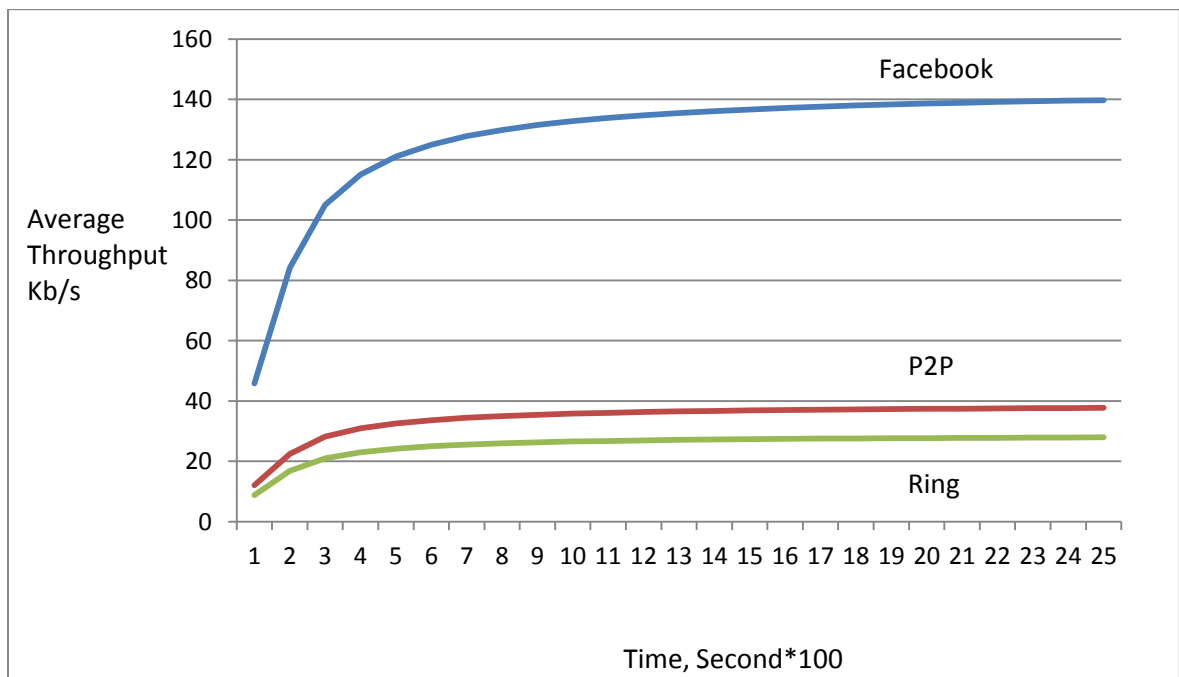


Figure 19-Trace1-Short video streaming with constant size of 10MB

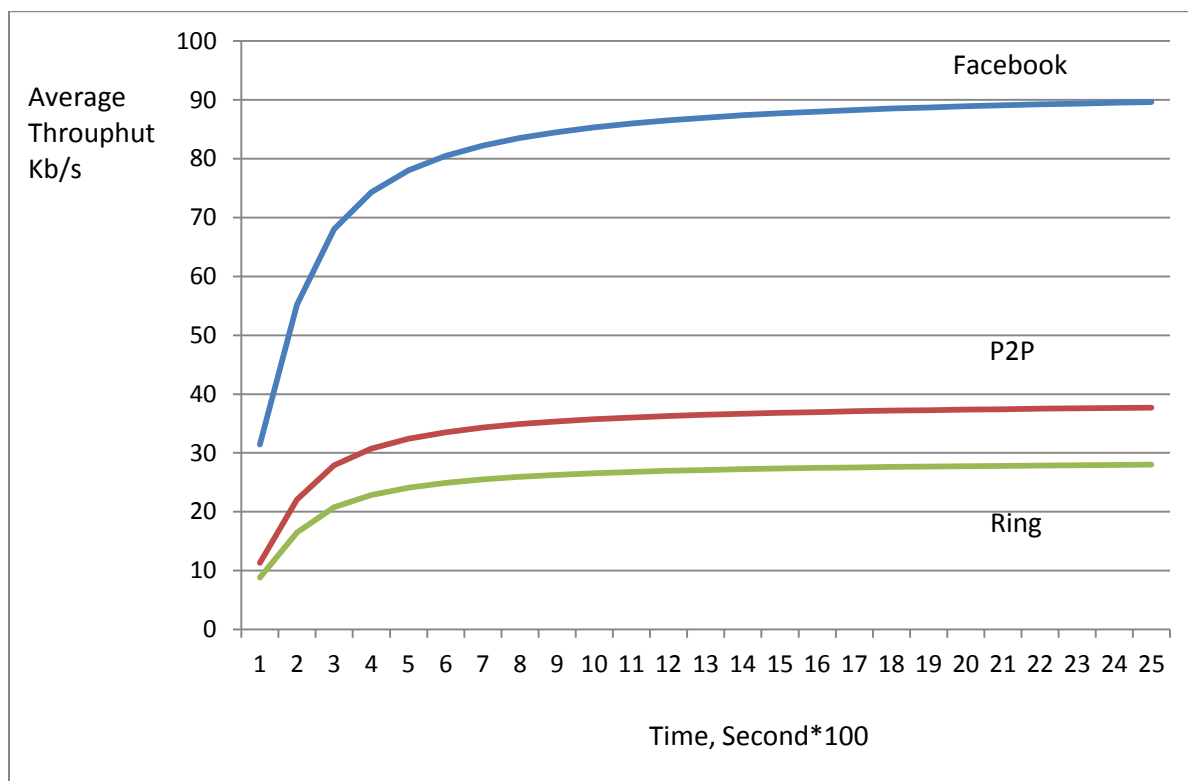




**Figure 20-Trace2-Short video streaming with constant size of 10MB**



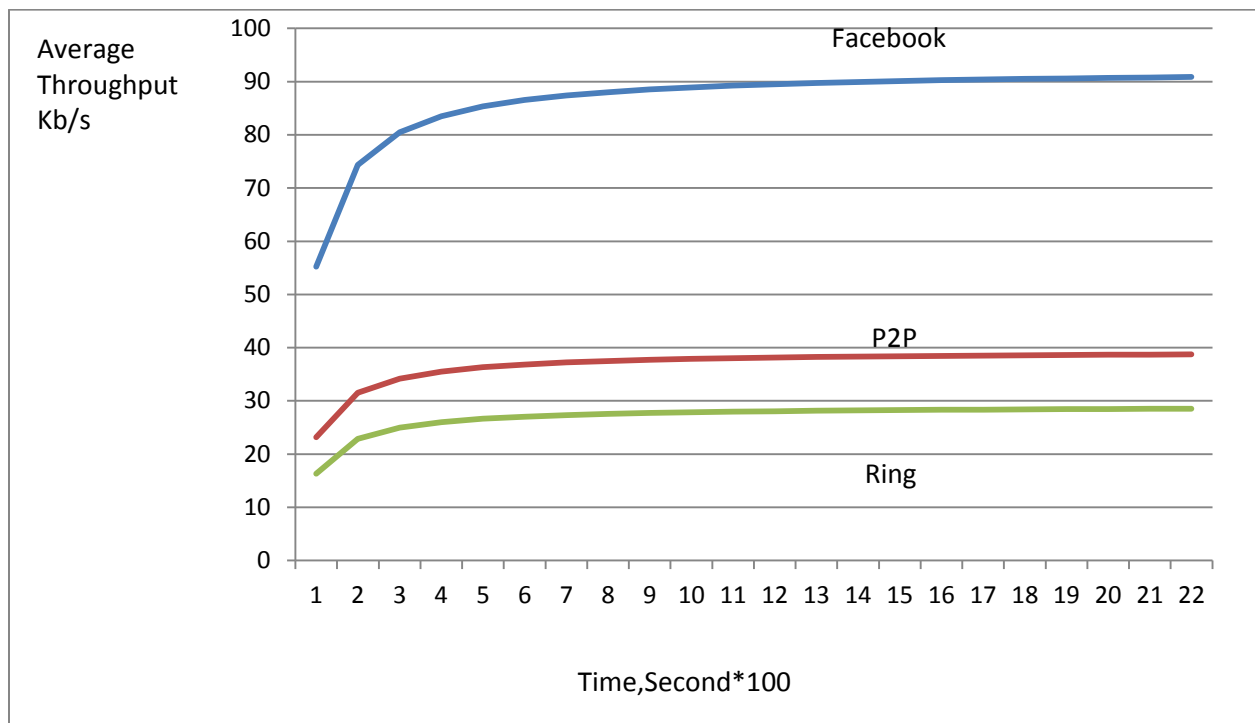
**Figure 21-Trace3-Short video streaming with constant size of 10MB**



**Figure 22-Trace4-Short video streaming with constant size of 10MB**

### 4.1.3 Short Video Streaming with Fixed Size of 50 MB

In 2009 YouTube started providing short videos in high definition, therefore making the size of the movies larger. Here we conducted an experiment based on a file size of 50 MB for short-video sharing, consequently creating a more realistic experiment.



**Figure 23-Trace4-Short video streaming with constant size of 50MB**

#### 4.1.4 Short video streaming with variable size

In this set of experiments we used a synthetic workload generator for simulating YouTube video parameters [66] in order to simulate short-video sharing in these three networks.

In [66], during five months of research, statistical behaviors of 250,000 popular and regular videos were analyzed. Soraya found that the average sizes of the most popular videos on YouTube are 10.6 MB, and the distribution of video file sizes is a Weibull distribution.

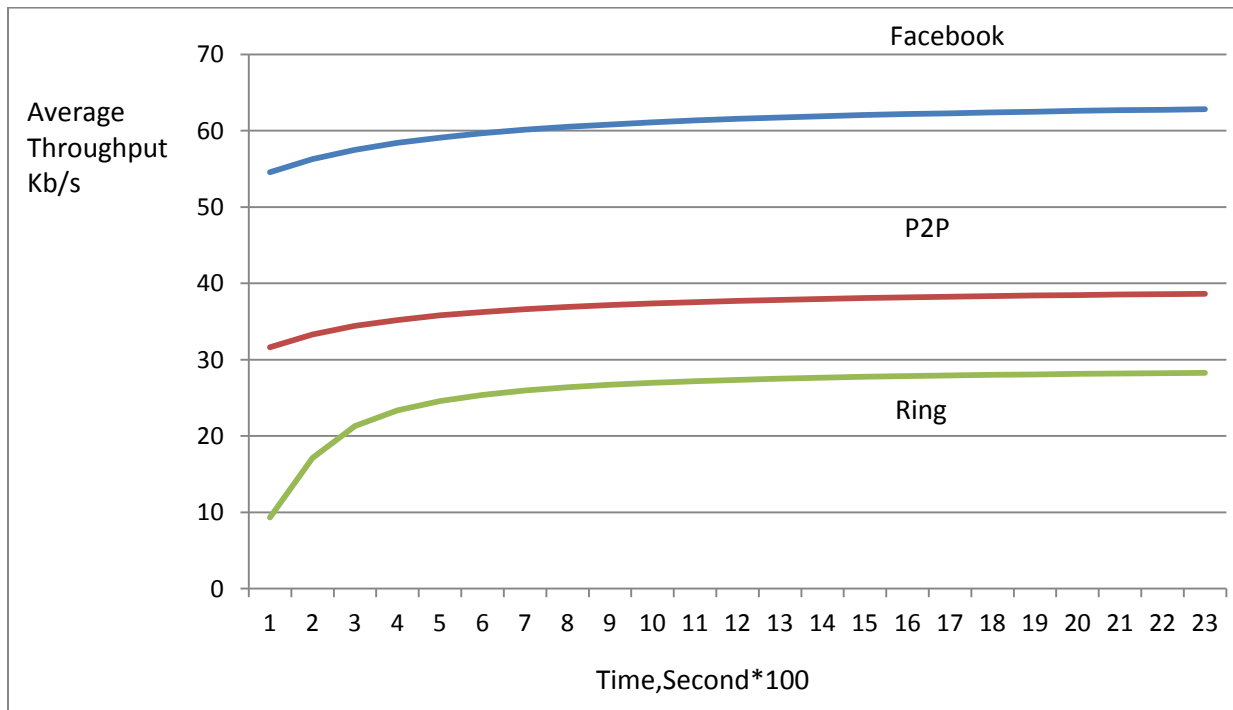
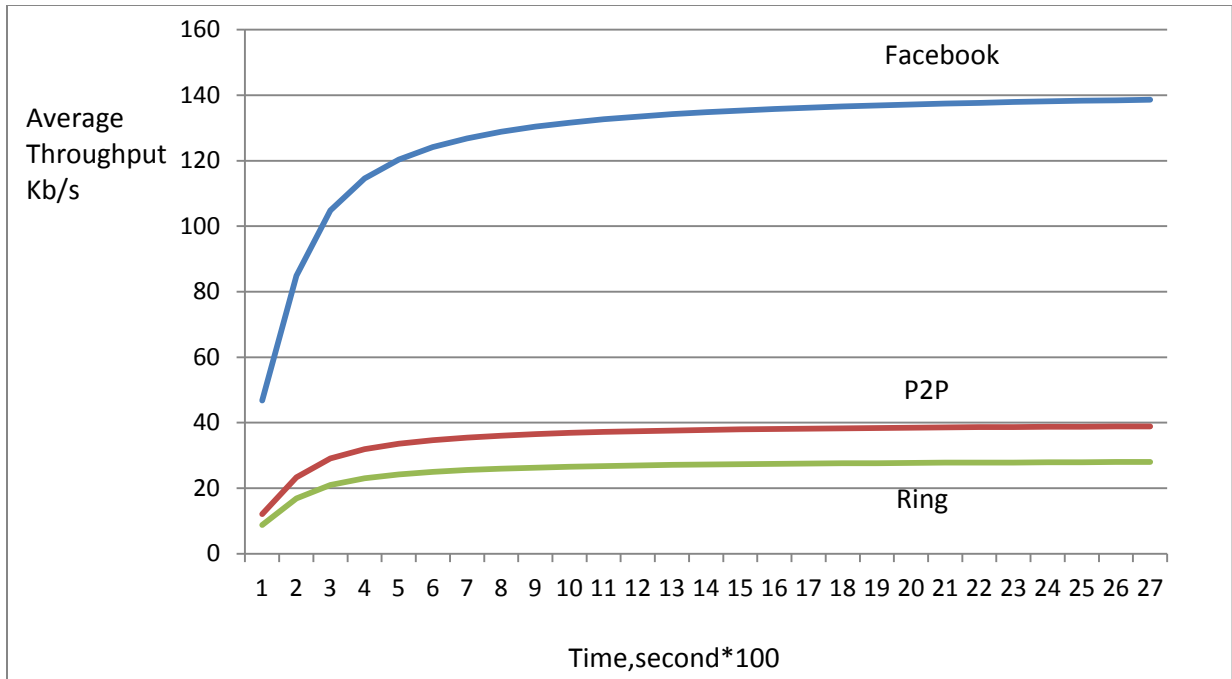
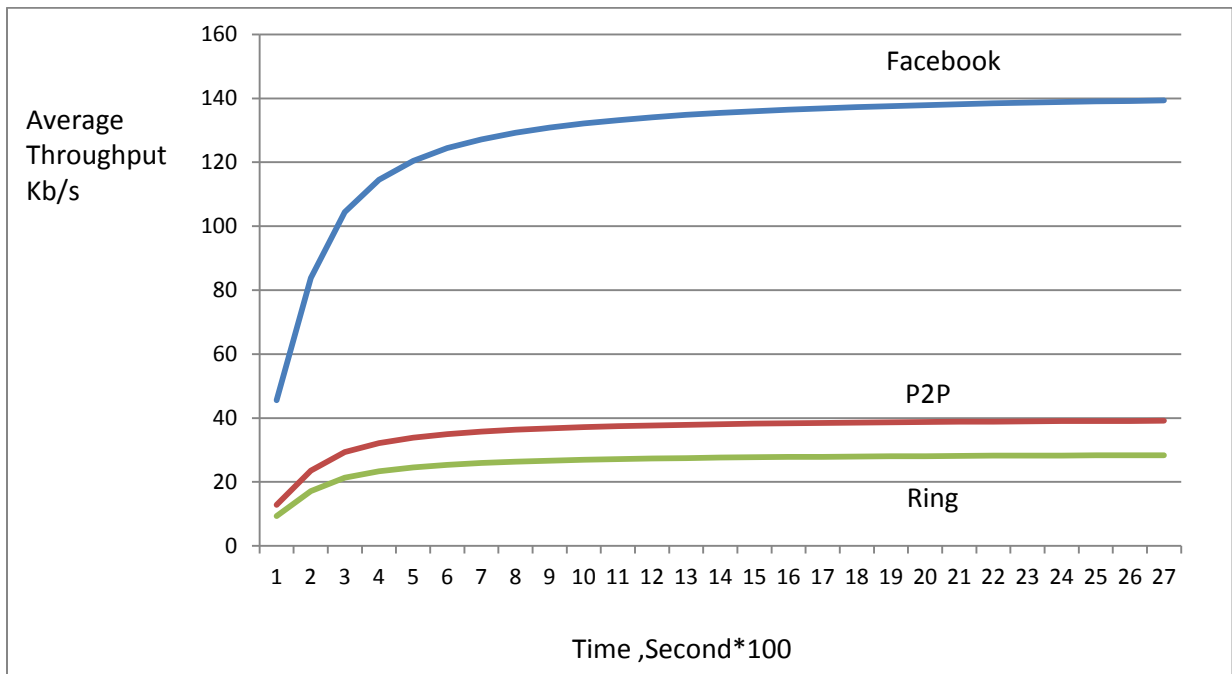


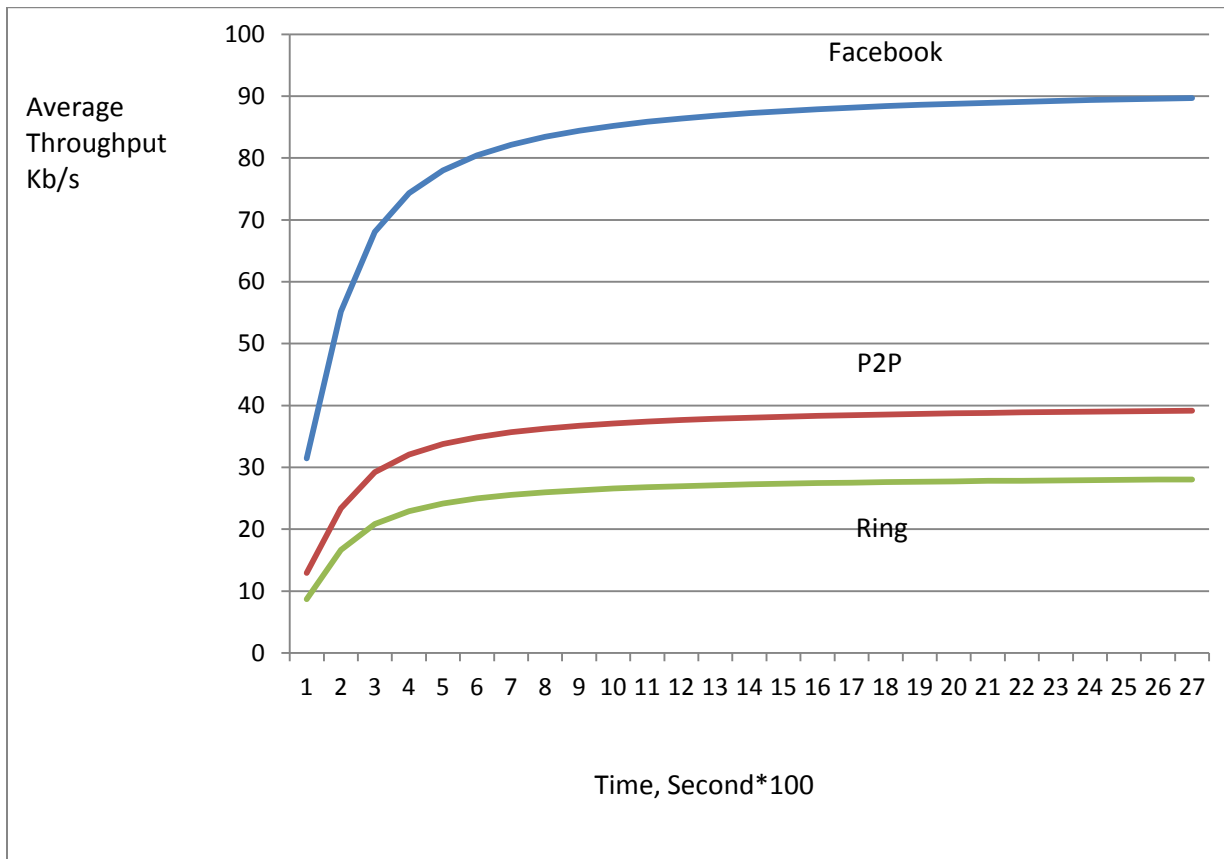
Figure 24-Trace 1-Short-video streaming with Weibull distribution of file sizes



**Figure 25-Trace 2-Short-video streaming with Weibull distribution of file sizes**



**Figure 26-Trace3-Short video streaming with Weibull distribution of file sizes**



**Figure 27-Trace 4-Short-video streaming with Weibull distribution of file sizes**

## 4.2 Discussion

Increasing the throughput out of the available bandwidth (2 MB) for each node will improve the quality of multimedia streaming. For instance in Trace 2, each node of Facebook network uses 6.76% of the bandwidth, while in unstructured P2P network each node uses 1.86% and structured P2P uses 1.38%. That means the download time and quality of multimedia streaming for each Facebook node is faster than the others. Tables 2, 3, 4 and figure 28 show that the download time is reduced in video streaming with Facebook network.

**Table 2-Percentage of bandwidth usage out of effective bandwidth with fixed size of 10 MB file sizes**

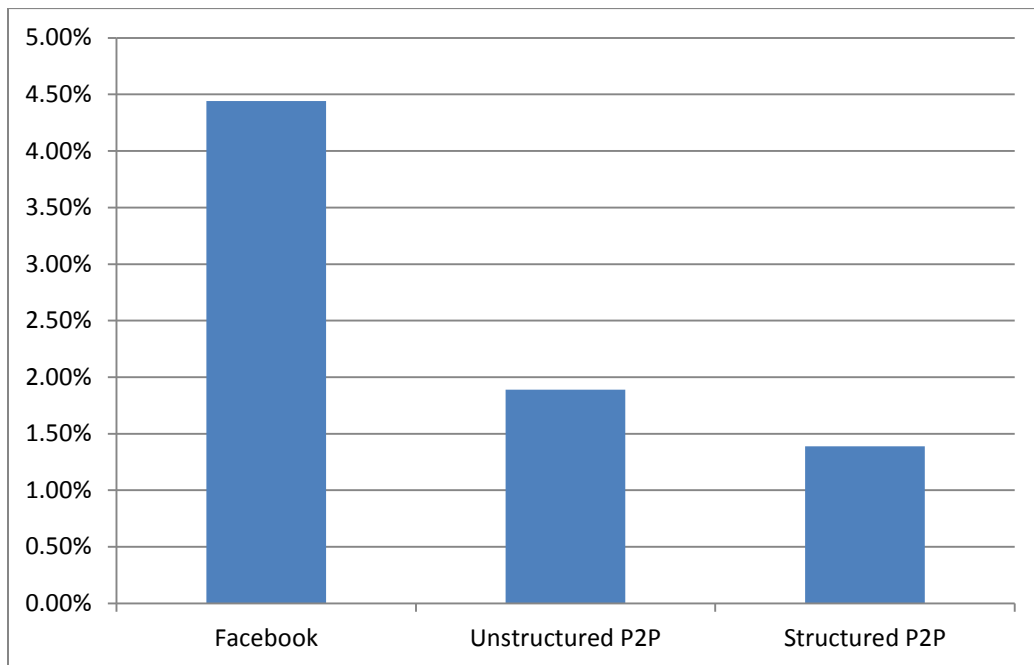
	Trace 1	Trace 2	Trace 3	Trace 4
Facebook	3.07%	6.76%	6.83%	4.38%
Unstructured P2P	1.85%	1.86%	1.84%	1.84%
Structured P2P	1.36%	1.38%	1.36%	1.36%

**Table 3-Percentage of bandwidth usage out of effective bandwidth with Weibull distribution file sizes**

	Trace 1	Trace 2	Trace 3	Trace 4
Facebook	3.07%	6.77%	6.81%	4.38%
Unstructured P2P	1.88%	1.90%	1.91%	1.91%
Structured P2P	1.38%	1.36%	1.38%	1.37%

**Table 4-Percentage of bandwidth usage out of effective bandwidth with 50 MB files sizes**

	Trace 4
Facebook	4.44%
Unstructured P2P	1.89%
Structured P2P	1.39%



**Figure 28-Percentage of bandwidth usage out of effective bandwidth with 50 MB**



### 4.3 Summary

In this chapter, the extensive simulations and experiments performed provide results that demonstrate the effectiveness of the Facebook network for short-video distribution and the P2P structured for the speed of information transfer.

All the experiments are tested on four different data sets. Facebook network shows 59.85% average improvement over unstructured P2P in short-video streaming with variable file size, and 70.93% average improvement over structured P2P. With a fixed-size file of 10 MB, Facebook has 70% average improvement over unstructured P2P and 60.92% average improvement over structured P2P. While extracting speed of information transfer, structured P2P had 16% average improvement over unstructured P2P and 60% average improvement over the Facebook network.

# Chapter 5

## 5. Conclusions and Future Works

This chapter of the thesis presents the conclusions of our research (Section 5.1), and Section 5.2 presents future works.

### 5.1 Conclusion

In 2011, the YouTube web site, which is a video sharing website, changed its user's interface, so that it began to function more like a social networking web site than it had previously done. When users in Google plus shared a multimedia file, other users would be able to see that video file on YouTube without logging into Google plus.

Researchers need to consider the impact of social-networking phenomena. By social-networking phenomena, we are referring to applications that have changed the way we communicate with each other, such as Facebook, Twitter, YouTube and Spotify. At this point in time computer programmers should consider incorporating different aspects of social networking into their design of web-based applications.

The Facebook network is based on client-server architecture, where users need to add each other and become friends before exchanging any information. Therefore a user may retrieve a video file included in the profile of another user which is stored in the Facebook server.

In this thesis a simulated environment is created in order to forecast the outcome of systems without actually building the systems themselves. In our simulation of a Facebook web site, we assume that all users are located on a P2P overlay network, and they are all online. This specific assumption will develop a fair comparison scenario for the simulations performed in this

work, since the Facebook user network graph is compared with other P2P networks. In P2P networks, while users are downloading and uploading files, they can exchange comments with each other without knowing one another. Therefore, the diversity of the network is higher in P2P networks as compared to the Facebook network. In the first experiment, this is the reason behind the higher speed of information transfer among nodes in P2P networks compare to Facebook. On average, the structured P2P network has a 16% faster rate of information transfer over the unstructured P2P network, and a 60% faster rate of information transfer over Facebook network. Thus, we conclude that in the structured P2P network, information can be transferred faster.

In the second group of experiments, we focused on multimedia sharing. The network that has the highest average throughput is the best choice for media sharing. The network stays stable for a considerable period of time, therefore throughput can describe how robust the network becomes for the purpose of sharing short video files. Furthermore, increasing the throughput out of the available bandwidth for each node will improve the overall quality of multimedia streaming, since the download speed for each user is increased. While considering media sharing by simulating short-video sharing, the Facebook network shows 61% average improvement in throughput compared to the P2P network, and 71% average improvement over ring network.

In the third experiment, when we test with file size of 10 MB, Facebook has 62% average improvement in throughput for media sharing over P2P, and 72% average improvement over the ring network. Thus Facebook is the most suitable network for sharing multimedia files between users since it has the highest average throughput.

## 5.2 Future Work

Previous works on creating a pure P2P web-based application failed due to their single point of failure such as, Napster. Analyzing social aspects of a Facebook-like network that has client-server architecture and modeling P2P networks gave us an idea for creating a P2P web-based application that has the characteristic of the social networking web sites. This is the main future work of this thesis.

Spotify is the first commercial attempt at using P2P architecture that takes advantage of social networking. [74] shows that the lack of privacy in OSN websites can be substituted by the P2P storage system, where users are in charge of their own data. This will be another advantage of using P2P with OSN-like networks. Studying other advantages and disadvantages of such system creates directions for future work of this thesis.

Most research in the field of P2P multimedia streaming and file-sharing applications focuses on creating modeling and simulation design. There has been a lack of research in statistical works in the field of multimedia sharing within overlay networks. Analysis of user behaviors in a P2P system supports the realistic implementation of the P2P system. An issue such as which random distribution describes users who are downloading popular videos on a P2P network is worthy of future research.

## 6. References

- [1] M. Gneiser, J. Heidemann, A. Landherr, M. Klier, F. Probst, “Valuation of online social networks taking into account users interconnectedness,” *Information Systems and e-Business Management*, pp. 61-84, 2010.
- [2] B. Alexander and A. Levine, “Web 2.0 Storytelling, Emergence of a New Genre,” in *educasue Review*. Vol. 43, no. 6, pp. 40-56, Nov/Dec, 2008.
- [3] <http://www.wiki.com/whatiswiki.htm>, Retrieved on 2/13/2012.
- [4] Official Web site of Napster- <http://www.napster.com/>, Retrieved on 3/11/2011
- [5] Official Web site of Spotify- <http://www.spotify.com/int/>, Retrieved on 3/11/2011
- [6] Official Web site of iTunes- <http://www.apple.com/itunes/>, Retrieved on 3/11/2011
- [7] X. Cheng and J. Liu, “NetTube: Exploring Social Networks for Peer-to-Peer Short Video Sharing,” in *IEEE Infocom the 28<sup>th</sup> conference on computer communication*, pp. 1152-1160, April 2009.
- [8] C. Xu, G. M Muntean, E. Fallon, and A. Hanley, “Distributed Storage-Assisted Data-Driven Overlay Network for P2P VoD Services,” in *IEEE Transactions on Broadcasting*, vol. 55, no. 1, pp. 1-10, March 2009.
- [9] Y. R. Choe, L. D. Schuff, M. J. Dyaberi, V. S. Pai. “Improving VoD Server Efficiency with BitTorrent,” in *Proceedings of the 15th International Conference on Multimedia (MULTIMEDIA’07)*, pp. 117-126, 2007.
- [10] F. Pakzad and A. Abhari, “Characterization of User Networks in Facebook,” in *the Proceedings of 13th Communications and Networking Simulation Symposium (CSN10) of SCS/SpringSim '10 in Collaboration with ACM/SIGSIM*, Orlando, FL, US, April 2010.

- [11] M.Gjoka, M. Sirivianos, A. Markopoulou, X. Yang, “Poking Facebook: Characterization of OSN Application,” in *Proceedings of Workshop on Online Social networks, Seattle, WA, USA. August 2008.*
- [12] Official Web site of Google Plus-<https://plus.google.com/>. Retrieved on 8/11/2011.
- [13] Official Web site of LinkedIn- <http://www.linkedin.com>. Retrieved on 8/12/2011.
- [14] Official Web site of Orkut- Online Social Network Web site, <http://www.orkut.com>, retrieved on 8/12/2011.
- [15] Official Web site of Twitter- <http://twitter.com/>. Retrieved on 8/12/2011
- [16] S. Milgram, “The Small World Problem,” *Psychology Today*, vol.2, no. 1, pp. 60-67, 1967.
- [17] J. .M Kleinberg, “The small-world phenomenon: An algorithmic perspective”, in *Proceedings of the 32nd ACM Symposium on Theory of Computing Annual ACM Symposium on Theory of Computing archive: Proceedings of the thirty-second annual ACM symposium on Theory of computing*, Portland, Oregon, 163-170.2000.
- [18] [http://www.canicashopping.com/2011\\_11\\_22\\_archive.html](http://www.canicashopping.com/2011_11_22_archive.html), Retrieved on 2/09/2012.
- [19] G. Kreitz and F. Niemelä, “Spotify – large scale, low latency, P2P music on-demand streaming,” in *Peer-to-Peer Computing. IEEE*, pp. 1–10. August 2010.
- [20] M. Goldmann and G. Kreitz, “Measurements on the Spotify Peer-Assisted Music-on-Demand Streaming System,” in *Proceedings. IEEE P2P*, 2011.
- [21] P. Gill, M. Arlitt, Z. Li, and A. Mahanti, “YouTube Traffic Characterization: A View From the Edge,” in *Proceedings of the 7th ACM Sigcomm Conference on Internet Measurement*, pp. 15-28, San Diego, USA, October 2007.

- [22] Q.Tang, B.Gu, A.Whinston, " Content Contribution in Social Media: The Case of YouTube," *45th International Conference on System Sciences*, Hawaii,pp. 4476-448. January 2012.
- [23] Overlay Networks, I. Stocia, <http://www.cs.virginia.edu/~cs757/slidespdf/757-09-overlay.pdf>, retrieved on 9/10/2011.
- [24] J. Jannotti, D. Gifford, K. Johnson, M. Kaashoek, and J. O'Toole, "Overcast: Reliable multicasting with an overlay network," *presented at Fourth Symposium on Operating System Design and Implementation (OSDI)*, San Diego, CA, October 2000.
- [25] T. Ho, M. Medard, J. Shi, M. Effros, and D. Karger, "On Randomized Network Coding," in *Proceedings of Allerton Conference on Communication, Control, and Computing*, October 2003.
- [26] S. A. Theotokis and D. Spinellis, "A survey of peer-to-peer content distribution technologies," *ACM Comput. Surv.*, vol. 36, no. 4, pp. 335–371, 2004.
- [27] "What Is P2P ... And What Isn't" Retrieved on 1/09/12.  
<http://openp2p.com/pub/a/p2p/2000/11/24/shirky1-whatisp2p.html?page=2>.
- [28] G, Coulouris, J, Dollimore and T.Kindberg. 2005. *Distributed Systems Concepts and Design*, 4rd ed. Edinburgh: Addison-Wesley, pp. 397-399.
- [29] A, S. Tanenbaum and M. V. Steen, *Distributed Systems Concepts and Paradigms*. 2007. 2nd ed. Upper Saddle River: Pearson Prentice Hall, pp. 43-53.
- [30] Y. Desmedt and J. J. Quisquater, "Public-key systems based on the difficulty of tampering (Is there a difference between DES and RSA?)," in *Proceedings on Advances in cryptology-crypto '86*, London, UK, pp. 111-117, Springer-Verlag LNCS 263, 1987.

- [31] Q. Lv, P. Cao, E. Cohen, K. Li, and S. Shenker, "Search and replication in unstructured peer-to-peer networks," in *Proceedings of the 16th international conference on Supercomputing*, (New York, NY, USA), pp. 84-95, ACM, 2002.
- [32] Gnutella-<http://rfc-gnutella.sourceforge.net/>. Retrieved on 2/11/2011.
- [33] S. Saroiu, K. P. Gummadi, and S. D. Gribble, "Measuring and analyzing the characteristics of Napster and Gnutella hosts," *Multimedia Systems Journal*, vol. 9, no. 2, pp. 170–184, 2003.
- [34] I. Stoica, R. Morris, D. Karger, M. F. Kaashoek, and H. Balakrishnan. "Chord: A scalable peer-to-peer lookup service for Internet applications," in *Proceedings. ACM SIGCOMM'01*, San Diego, CA, August 2001.
- [35] D. karger, E. Lehman, F. Leighton, F. Levine, M. Lewin, R. Panigrahy "Consistent hashing and random trees: Distributed caching protocols for relieving hot spots on the World Wide Web," in *Proceedings of the 29th Annual ACM Symposium on Theory of Computing*, El Paso, TX, , pp. 654–663, May 1997.
- [36] Official Web site of PeerSim- <http://peersim.sourceforge.net/>, Retrieved on 3/09/2011.
- [37] Official Web site of GPS- <http://www.cs.binghamton.edu/~wyang/> Retrieved on 3/09/2011.
- [38] I. Baumgart, B. Heep, and S. Krause. 2007. "OverSim: A Flexible Overlay Network Simulation Framework," in *Proceedings of 10th IEEE Global Internet Symposium*, pp.79-84, May 2007.
- [39] Official Web site of OverSim-<http://www.oversim.org/>, Retrieved on 15/11/2011.
- [40] B. Cohen. BitTorrent protocol specification.  
<http://www.bitconjurer.org/BitTorrent/protocol.html>, Retrieved on 13/11/2011.



- [41] B. Cohen, “Incentive Build Robustness in BitTorrent,” in *Proceedings of the First Workshop on the Economics of Peer-to-Peer Systems*, Berkeley, May 2003.
- [42] A. Legout, G. Urvoy-Keller, and P. Michiardi. “Rarest First and Choke Algorithms Are enough,” *ACM IMC*, Rio de Janeiro, Brazil, October 2006.
- [43] J. Li, Y. Cui, and B. Chang, “Peer Streaming: Design and Implementation of an On-Demand Distributed Streaming System with DRM Capabilities,” *Multimedia Systems Journal*. 2007.
- [44] N. Liogkas, R. Nelson, E. Kohler and L. Zhang. “Exploiting BitTorrent for fun (but not profit),” in *Proceedings. International Workshop on Peer-To-Peer Systems*, 2006.
- [45] M. Castro , P. Druschel , A. M. Kermarrec and A. I. T Rowstron, “Scribe: A Large-Scale and Decentralized Application-Level Multicast Infrastructure,” *IEEE Journal of Selected Areas in Communications*, Vol. 20, No. 8, pp. 1489–1499, 2002.
- [46] S. Ratnasamy, P. Francis, M. Handley, R. Karp and S. Shenker, “A Scalable Content-Addressable Network,” in *Proceedings of the ACM SIGCOMM* , pp. 161–172, San Diego, USA, 2001.
- [47] D. Constantinescu, “Overlay Multicast Networks: Elements, Architectures and Performance,” Department of Telecommunication Systems School of Engineering Blekinge Institute of Technology, Karlskrona, Sweden, PHD Thesis, No 2007:17, 2007.
- [48] D. Erman, “On BitTorrent Media Distribution,” Department of Telecommunication Systems School of Engineering Blekinge Institute of Technology, PHD Thesis, Karlskrona, Sweden, No 2008:04, 2008.

- [49] J. Wang, W. Yurcik, Y. Yang, and J. Hester, "Multi-Ring Techniques for Scalable Battlespace Group Communications," *IEEE Communications Magazine*, Vol. 43, No. 11, pp. 124–133, November 2005.
- [50] A. Matrawi, and I. Lambadaris, "A Survey of Congestion Control Schemes for Multicast Video Applications," *IEEE Communications Surveys and Tutorials*, Vol. 5, No. 2, pp. 22-31, December 2003.
- [51] E. Altman and T. Jimenez, "NS Simulator for beginners," University de Los Andes, Merida, Venezuela and Essi, Sophia-Antipolis, France. 2003-2004
- [52] K. Fall. "The ns Manual (formerly ns Notes and Documentation)," 2011.
- [53] F. James, Kurose and Keith W. Ross, "Computer Networking: A Top-Down Approach Featuring the Internet," (3rd ed.), USA: Addison-Wesley, 2004.
- [54] M. B. Tajali, "A Peer-to-Peer Delivery System for Internet Short Video Sharing," Master of Applied Science in the Program of Electrical and Engineering, Toronto, Canada, 2010.
- [55] M. Li, M. Claypool, R. Kinicki, and J. Nichols, " Characteristics of Streaming Media Stored on the Web," in *ACM Transactions on Internet Technology*, Volume 5, Number 4, pp. 601-626, November 2005.
- [56] S. Jin, A. Bestavros, "GISMO: a Generator of Internet Streaming Media Objects and workloads," in *ACM SIGMETRICS Performance Evaluation Review*, November 2001.
- [57] W. Tang, Y. Fu, L. Cherkasova, A. Vahdat, "MediSyn: a synthetic streaming media service workload generator," in *Proceedings of the 13th International Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV)*, CA, USA, June 2003.

- [58] C. Huang, J. Li, and K. W. Ross, "Can Internet Video-on-Demand be Profitable? ," in *ACM SIGCOMM Computer Communication Review*, Volume 37, Number 4, pp.133-144, October 2007.
- [59] M. Halvey and M. Keane, "Exploring Social Dynamics in Online Media Sharing ", in *Proceedings of the 16th international conference on World Wide Web, Banff, Canada*, May 2007.
- [60] A. Mislove, M. Marcon, K. Gummadi, P. Dreschel, B. Bhattacharjee, "Measurement and Analysis of Online Social Networks," in *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement Internet Measurement*, pp. 29 - 42, San Diego, California, USA, October 2007.
- [61] J. Paolillo, "Structure and Network in the YouTube Core," in *Proceeding of the 41st annual Hawaii International Conference on System Sciences (HICSS'08)*, pp.156, 2008.
- [62] M. Zink, K. Suh, and J. Kurose, Watch Global, "Cache Local: YouTube Network Traffic at a Campus Network - Measurement and Implications," in *Proceedings of ACM/SPIE MMCN '08 conference*, Volume 6818, pp. 5-13 San Jose, USA, January 2008.
- [63] Cha, H. Kwak, P. Rodriguez, Y. Ahn, and S. Moon, "I Tube, You Tube, Everybody Tubes: Analyzing the World's Largest User Generated Content Video System," in *Proceedings of the 7th ACM SIGCOMM conference on Internet Measurement*, pp.1-14, San Diego, USA, October 2007.
- [64] Daum UCC. <http://ucc.daum.net>.
- [65] X. Cheng, C. Dale, and J. Liu, "Understanding the Characteristics of Internet Short Video Sharing: YouTube as a Case Study", *Technical Report arxiv: 0707.3670v1 {cs.NI}*, Cornell University, arxiv e-prints, July 2007.

- [66] A. Abhari, M. Soraya, “Workload Generation for YouTube,” *Multimedia Tools and Applications journal*, Vol 46, pp. 91-118, June 2010.
- [67] P. Garbacki, D. Epema, J. Pouwelse, and M. V. Steen, “Offloading Servers with Collaborative Video on Demand,” in *Proceedings of the 7th International Workshop on Peer-to-Peer Systems (IPTPS'08)*, February 2008.
- [68] C. Huang, J. Li, and K. Ross, “Can Internet Video-on-Demand be Profitable?,” in *Proceedings of ACM SIGCOMM'07*, January 2007.
- [69] J. D. Mol, J. Pouwelse, M. Meulpolder, D. Epema, and H Sips, “Give-to-Get: Free-riding-resilient Video-on-Demand in P2P Systems,” in *Proceedings of the 15th SPIE/ACM Multimedia Computing and Networking (MMCN'08)*, 2008.
- [70] J. Postel, “User Datagram Protocol,” RFC 768: pp.1-3, 28 August 1980.
- [71] C. Wang, B. Li, “Peer-to-peer overlay networks: A survey”, *Technical Report, Department of Computer Science, HKUST*, February 2003.
- [72] A. Stuart, K. Ord, S. Arnold, “Classical Inference and the Linear Model. Kendall's Advanced Theory of Statistics,” Sixth edition, pp. 25.37-25.43, London: Arnold, 1999.
- [73] S. Talaei and A. Abhari, “Which type of Overlay Network Performs Better for Multimedia distribution and Information sharing,” in *the Proceedings of 15th Communications and Networking Simulation Symposium (CSN12) of SCS/SpringSim '15 in Collaboration with ACM/SIGSIM*, Orlando, FL, US, April 2010.
- [74] D. Schioberg, “A peer-to-peer infrastructure for social networks,” Master Thesis, TU Berlin, Berlin, Germany, pp.3-47, December 17, 2008.
- [75] <http://www.cosbi.eu/index.php/research/prototypes/graph>, Retrieved on 2/6/2012.